现代外国统计学优秀著作译丛

统计学

STATISTICS

[美] David Freedman
Robert Pisani
Roger Purves
Ani Adhikari 著

魏宗舒 施锡铨
林举干 李 毅 译
吕乃刚 范正绮

吴喜之 校

中国统计出版社

现代外国统计学优秀著作译丛

统 计 学

（第二版）

〔美〕David Freedman
Robert Pisani
Roger Purves
Ani Adhikari
著

魏宗舒 施锡铨
林举干 李 毅
吕乃刚 范正绮
译

吴喜之
校

中国统计出版社

(京)新登字 041 号

图书在版编目(CIP)数据

统计学/(美)Freedman,D. 等著;魏宗舒等译.
—北京:中国统计出版社,1997.1
(现代外国统计学优秀著作译丛)
ISBN 7-5037-1966-4

Ⅰ.统…
Ⅱ.①F… ②魏…
Ⅲ.统计学—教材
Ⅳ.C8

中国版本图书馆 CIP 数据核字(95)第 11405 号

著作权合同登记:图字:01-97-0001 号

中国统计出版社出版
(北京三里河月坛南街 75 号 100826)
新 华 书 店 经 销
科伦克三莱印务(北京)有限公司印刷
*
850×1168 毫米 32 开本 23.75 印张 61 万字
1997 年 10 月第 1 版 1997 年 10 月北京第 1 次印刷
印数:1—4000 册
*
定价:49.00 元

版权公告：

Copyright notice：

统计学

STATISTICS

(Second Edition)

[美] David Freedman

Robert Pisani

Roger Purves

Ani Adhikari

现代外国统计学优秀著作译丛
专家委员会

出 版 说 明

为了加强对国外统计理论与实践的研究和了解，全面反映国外统计科研和教学的发展，促进我国统计教学改革和教材内容更新，在国家统计局领导的大力支持下，全国统计教材编审委员会组织翻译出版了这套“现代外国统计学优秀著作译丛”。

随着我国社会主义市场经济体系的逐步建立，统计教育正面临着十分严峻的挑战。一方面，在社会主义市场经济条件下，不论国家的宏观经济调控还是企业的生产经营管理，都要求准确地把握市场运行的态势，科学地分析经济中各种错综复杂的关系，因而，对统计信息的需求越来越大，对统计人才的业务素质提出了更高的要求；另一方面，我国过去的统计教育模式是按为高度集中的计划经济管理体制服务的要求建立的，培养的统计人才的知识结构比较单一，难以适应经济体制、统计体制改革的需要。为使统计人才的培养适应建立社会主义市场经济体制的需要，满足二十一世纪现代化建设的要求，缩小与国际先进水平的差距，基础在教育，关键在教材。在继续组织有关专家、学者编写一批反映国内统计科学和统计实践发展的新教材的同时，必须尽快引进并翻译出版一批外国先进统计教材。这是学习外国先进统计知识的一种直接而且十分有效的方式，对于推动国内统计教材内容更新和教学改革，造就一大批具有渊博知识和多方面

业务技能的复合型人才，具有十分重要的意义。

为了做好这套丛书的翻译出版工作，全国统计教材编审委员会成立了现代外国统计学优秀著作译丛专家委员会，对国外统计著作的出版和使用情况进行了调查研究，分析了国内对外国统计教材的需求，在此基础上制定了翻译著作选题规划。在这套丛书的翻译出版过程中，我们得到了国内外有关专家、有关院校统计系和国外有关出版公司的大力帮助和支持，在此表示衷心的谢意。

全国统计教材编审委员会
1995年7月

译　者　序

干了多年的统计教育，我接触过许多种不同类型的学生。在工科院校为本科生和研究生开设《工程数学》讲解概率论与数理统计，学生们关心的是如何将数据代入公式以完成工程计算。由于不存在适合众多统计问题的简单统计公式而使有些人觉得这门课程较难。后来为理科大学统计运筹系开设各类统计课程，理科的学生，尤其是研究生，常常关心数理统计的定理，其条件与结论以及可能的推广等等，较少关心所研究问题的统计背景，于是统计被纳入纯数学的学究式轨道。最近几年执教于财经大学，财经类学生对统计的认识与要求又有明显的不同，少数人误认为统计只是填写生产经营报表而不感兴趣，而大多数学生被原苏联的计划式统计与来自西方的统计互相交错重迭甚至相悖搞得晕头转向。在改革开放及社会主义市场经济逐步建立与完善的今天，如何引导学生定量地描述社会经济的现象与本质，换句话，如何上好《统计学原理》这门课程，这一问题，明明白白地摆在了统计教育工作者的面前。在这里，我们向大家推荐 David Freedman，Robert Pisani，Roger Purves 与 Ani Adhikari 的《统计学》(第二版)。作为教本或者参考书，它至少给予从事统计工作的人们在思想方法上许多启迪。

其实，不管是工科的理科的，还是文科的财经类的学生，只要他们接触较多数据，都面临着一个共同需要解决的问题：如何考虑一个统计问题。作为实验或观察的结果，一大堆包藏着万物奥秘的数据“乱七八槽”地呈现在人们的面前，有时使人感到无从入手，急切地找一个统计公式将数据代入常常会发生差错。此时最需要的是对数据有一定的“悟性”以及对形形式式的统计处理有直观的理

解。也就是说，人们应当具有一定的统计思想。《统计学》正是引导人们对统计问题进行思考的一本杰出的教材，它是迄今为止我们所见到的最出色的关于统计思想的入门书。它以清晰的日常语言和各种类型的例子向人们介绍解决统计问题的若干技巧，展示了统计在诸如经济、教育、遗传、医药、物理、化学、环境污染、政治及社会科学、心理学等方面所起的至关重要的作用。在整个展开过程中除了加减乘除等简单运算之外几乎不用什么数学语言，这为各行各业有兴趣接受统计思想的人创造了极大的方便。

本书的主干部分有29章。施锡铨翻译了第1—6章，第7—12章由李毅负责，吕乃刚译完了第13—18章，林举干完成了第19—23章和第27—29章的全部工作，第24—25章及第26章分别由范正绮与魏宗舒翻译。施锡铨与林举干对序、注释和习题答案进行了翻译并分别对全书进行了审校。最后由施锡铨对全书作了统一定稿。在翻译本书过程中，我们发现原书写得精彩紧凑，作者所用语言非常生动，比喻相当贴切，例子与习题精心选择，某些提法的用词值得人们对主题一再回味。我们确实增益不浅。如果读者从本书的译文很少有这种体会，那只能怪罪于译者的翻译水平，对此我们只能表示深深的歉意。

但愿本书能对各类统计工作者有所启发与帮助，尤其使社会经济、财经类的广大读者有所获益。

施锡铨

1995年3月于上海财经大学

献给　Jerzy Neyman（1894—1981）

Neyman 出生于俄罗斯，1938 年来到美国之前曾在波兰和英国工作。他是当代伟大的统计学家之一。

第一版序

Syrens 所唱的歌，或者 Achilles 藏身于女人中所用的假名，虽然使问题困惑费解，都没有超出一切猜想。

——Thomas Browne 爵士(英国,1605—1682)

给读者

我们将告诉你一些借助于统计方法已得到研究的有趣的问题，并且教你如何自己去使用某些统计方法，解释方法行得通的理由以及当其他人使用它们时所要注意的事项。对于大多数人来说数学记号似乎只会使问题混乱，因此我们将利用词句、图和表——几乎不用任何 x 或者 y 去做这件事。即使专业数学家阅读技术书时，无论他们如何认真，他们的眼睛也常常略过方程式。他们真正需要的是一个合意的朋友，这位朋友将解释概念和描绘藏身在方程式后面的图象。对于阅读本书的读者，我们试图成为那样的朋友。

统计学是什么？

统计学是对令人困惑费解的问题作出数字设想的艺术。

- 应该怎样设计实验来测定新药的疗效？
- 什么东西引起父母与孩子之间的相象，并且那种力量有多强？
- 通货膨胀率如何测定？失业率呢？它们怎样联系起来？

· 赌场为什么在轮盘赌上得益?

· 盖洛普民意测验怎么能够使用仅仅几千人的样本预测选举结果?

这些都是困难的问题,在分析它们的过程中统计方法有相当大的帮助。这些方法在数百年时期内经过一些致力于寻求艰难问题答案的人的努力而得到发展。其中的一些人在本书的以后章节中将会介绍。

概要

第一部分是关于适当地设计实验,使得从结果中可以得到有意义的结论。从医学、教育和城市政策中获得例子,并且对一些设计很糟糕的和误导的研究进行分析。对一些在判断一个设计的质量时需要询问的问题作了解释。大多数研究产生了如此多的数字以致于不可能全部吸收它们:必须进行概括。描述性统计是描述和概括数据的艺术。学科的这个分支在第二部分中介绍。在第三部分继续讨论,在那里焦点放在分析关系——例如,收入怎样依赖于教育。

许多统计推理依赖于第四部分中所讨论的概率论。通过机会模型建立它们之间的联系,关于机会模型在第五部分展开并在第七部分中应用以解决一些有关测量误差和遗传学的问题。第六部分和第八部分的统计推断,讲解了如何从样本去作有效的概括。第六部分论述估计理论并回答如下问题:盖洛普民意测验怎样取一组样本用来估计民众中民主党的百分数?这个估计相差可能有多远?利用显著性检验,第八部分讲述了怎样判断一组样本肯定还是否定它所来自的母体的有关假设。

现今,推断是统计学领域内的同行们最感兴趣的一个分支。但是,非统计学家常常发现描述性统计是更为有用的分支——而且是一种容易理解的分支。

学科的梗概在第1章至6章,16章至21章,23章和26章中

介绍。融会贯通这些东西之后，读者可以到处浏览，接下去要读的章节也许是8，10，和13章。

习题

每一章中编号的小节通常以一组习题结束，答案在本书的结尾。如果当这些习题出现时你进行练习并且对照书后面以检查答案，你将在新技巧方面获得实践——并且了解你掌握的程度。

除了第1章和第7章之外的每一章都以一组复习题结束；本书不给出这些习题的答案。在求解它们时，你可能想往回逐页地翻书直到发现有关的公式为止。但是，往回阅读本书将证明是非常令人沮丧的。复习题需要的知识远远超过公式。它们要求大致的猜测和定性的判断，而且对于发生的事需要一个很好的直观理解。发展那种理解的途径是向前阅读本书。

为什么本书包含了如此多的无法用简单化公式来解决的习题？理由是几乎没有什么实际生活中的统计问题可以用那种方式解决。盲目地引进统计公式引起过许多疑惑。因此本书教你一个不同的办法：思考。

感谢

写作得到了福特基金会(1973—1974)和加利福尼亚大学董事会(1974—1975)的资助。我们希望感谢 Earl Cheit 和 Sanford Elberg(加利福尼亚—伯克莱大学的两位)的帮助和在关键时刻的鼓励。

我们的打字员是 Julia Rubalcava。她始终超出常情地耐心与我们在一起并且相当灵巧地将我们的潦草书写转换为打字稿。我们也得益于 Peggy Darland 和 Beatrice Rizzolo，他们将混乱的事情管理得井井有条。

Dana Fradon 绘制了漫画，而 Dale Johnson 则绘制了技术图形。计算机图由 David Draper，Roberta Heintz，Richard Lockhart，

和 Rick Persons 操作。

Frank Anscombe (Yale), Diccon Bancroft (Yale), Leo Breiman (UCLA), Merrill Carlsmith (Stanford), Persi Diaconis (Stanford), David Lane (Minnesota), Richard Light (Harvard), Gerry Mendelsohn (California 大学, Berkeley), Tom Rothenberg (California 大学, Berkeley), Bruce Rothschild (UCLA), Amos Tversky (Jerusalem), 和 Geoff Watson (Princeton) 提供了对原稿的非常有帮助的评论。

数年内，我们教程中的成员对本书逐次的草稿作了成百个有用的评注。两个最有察觉能力的学生评论家是 Carole Rohwer 和 Elton Sherwin。

最后，我们要感谢我们的编辑，Donald Lamm。不知怎么一来，他把一些一直在发展着而逐渐形成的手稿变成为一本书。

第二版序

在统计学的第一版面世十年后，开始显示出一定的年代痕迹，比如，到了 1987 年，1977 年的价格和收入由于相差一个因子 α 而过时。其它的数据组也需要更新。我们决定对本书作少许修订，放进一些新的数据并改进某些我们很少喜欢的段落。然而，一事引出另一事，两年以后用较大范围的重写形式我们完成了此项工作。

论题，以及描述的一般次序保持大致相同；因此也保持了教科书的味道。采用很少重复的材料我们尽量使得讲述精练些。这样为许多新的例子和习题留出了篇幅。第 2 章关于观察研究和第 9 章关于相关是完全重写的，而且描述更加集中。关于概率论的第 13 章和 14 章经过改编；现在讨论了条件概率，并且去除了非标准的术语。第 27 章关于两样本 Z—检验被延伸到覆盖随机化试验的内容。

第二版在 Berkeley，Stanford 和犹太州课堂试用过。我们认为它具有第一版的同样的全部优点，而且比起第一版来有较少的缺点。

关于第二版的致谢。

再一次地，由 Dana Fradon 绘制漫画以及由 Dale Johnson 绘制技术图型。我们感谢 Richard 和 Adele Cutler，Charles Everett，Mark Hansen，以及 Othmar Pfannes 在计算机上的奉献工作。图象在附录中被编成电码。

电码由 A. A. E. C. ,Pennington ,N. J. 设计。印刷由 Waldman 制图社,Pennsauken ,N. J. ,和 Berkeley 的排字车间负责。

有益的评注和建议来自于许多方面。我们特别要感谢 John Cairns, Persi Diaconis, Fred Katz, Peter McCullagh, Ludolf Meester, Lincoln Moses, Bill Navidi, Diana Petitti, Jaimie Robins, Bernie Saffran, Julie Shaffer, Shanna Swan, Amos Tversky,和 Hans Zeiseld。最后,我们向第一版——和第二版许多草稿的众多学生与读者表示我们的感谢。

目　　录

第三部分　相关与回归

第四部分 概 率

第五部分　机会变异

第六部分　抽　　样

第七部分　机会模型

第八部分　显著性检验

第一部分　实验设计

1
对照实验

总是做得对。将使一些人满意，而使其余的人惊讶。

——马克·吐温(美国 1835—1910)

1. 脊髓灰质炎疫苗的现场试验

一种新药问世。怎样设计一个试验来测试它的效果呢？最基本的办法就是比较[1]。把药分给处理组的病人，而把其他病人作为对照——他们不接受治疗。然后比较两个组的反应。病人将以随机的方式被分到处理组或对照组，整个实验在双盲情况下进行。所谓双盲是说：不管是病人，还是测定反应的医生都不应知道谁在处理组谁在对照组。这些想法应在一次真实的现场试验中逐步展开[2]。

1916 年第一次脊髓灰质炎流行病袭击了美国，此后的四十年期间，造成了成千上万受害者，尤其是儿童。到了二十世纪五十年代，发现了一些预防这种病的疫苗。由 Jonas Salk 培育的一种疫苗似乎最有希望。在实验室试验中，它被证实是安全的并且能产生针对脊髓灰质炎的抗体。需要一次大规模的现场试验以观察在实验

室之外这种疫苗是否能保护儿童防御脊髓灰质炎。

1954 年美国公共卫生总署决定组织这样一类实验。实验对象是那些处于最易受袭击的年龄组——1、2、3 年级的儿童。现场试验选在遍布全国的一批学区里进行，而这些学区据信脊髓灰质炎的危险最为严重。二百万儿童牵扯进这次试验，其中五十万接种了疫苗。一百万儿童故意不给予接种疫苗而另五十万儿童则拒绝了接种。

这就说明了对照方法。只有处理组中的对象接种疫苗，其余对象不进行处理而被用作对照。然后比较两组反应以发现处理是否产生了差异。在脊髓灰质炎疫苗现场试验中，处理组与对照组具有不同的大小，但这无关紧要。调查人员比较两组间每十万人患脊髓灰质炎的比率。用观察比率来代替绝对数以调整两组在大小上的差异。

医德上的一个麻烦问题是，应不应该给所有儿童接种疫苗?一个答案是就新药而言，即使在广泛的实验室试验之后，仍有可能不清楚好处是否超过风险。需要现场试验以查清现实生活中使用的结果。当然，即使没有对照，给大量儿童接种疫苗似乎也可能提供确证。举例来说，假如 1954 年脊髓灰质炎的发病确比 1953 年急剧下降，似乎就可以证实脊髓灰质炎疫苗的效果。然而事实并不一定如此，因为脊髓灰质炎是一种每年发病率变化很大的流行病。1952 年大约有 60 000 个病例;而 1953 年仅有其一半。没有对照，1954 年的低发病率只能表明两件事中的一件:要么疫苗是有效的，要么当年没有流行此病。

要了解疫苗是否起作用，其唯一途径是使一部分儿童不接种疫苗。当然，只有取得父母同意的儿童才能接种疫苗。于是一个可行的方案是:那些取得父母赞同的儿童组成处理组并接种疫苗。其余的儿童组成对照组。众所周知高收入家庭的父母常常比低收入家庭的父母更赞成接种疫苗。这可能引起关于疫苗的偏性，因为高收入家庭的儿童更易遭受脊髓灰质炎的伤害。

这一点最初似乎是矛盾的，但是脊髓灰质炎是一种卫生方面的疾病。那些生活在较差卫生环境的儿童，在他们童年早期尚受到来自母亲的抗体保护时趋向于感染轻微的这种病症。经受感染以后，他们产生了自身的抗体以抵抗以后更厉害的传染。居住在良好卫生环境里的儿童不会产生这些抗体。

统计教育就是为了避免偏性而使处理组与对照组除了处理这一点之外应尽可能地相似。于是，两组反应之间的任何差异只能归因于处理而不是其它事因。如果两个组关于处理之外的某种因素有所不同，那么这个其他因素的效应可能同处理的效应混淆在一起。分离这些效应可能是困难的或者是不可能的。混淆是偏性的主要根源。

关于脊髓灰质炎疫苗现场试验，曾提出若干方案。小儿麻痹症全国基金会(NFIP)希望对所有小学 2 年级的并取得父母同意的儿童接种疫苗，而将 1 年级及 3 年级的儿童留作对照。这个 NFIP 方案为许多学区所接受。但是，脊髓灰质炎是一种通过接触传播的传染病。因此 2 年级的发病率有可能高于 1 或 3 年级的发病率。这将使研究的结果不利于疫苗。或者 2 年级的发病率可能比较低，使得研究的结果有利于疫苗。此外，处在需要得到父母同意的处理组的儿童肯定同那些处在不需要父母同意的对照组中的儿童有着不同的家庭背景。采用 NFIP 方案，处理组将包含过多的来自高收入家庭的儿童，使得这个组比起对照组来更易感染脊髓灰质炎。造成一个不利于疫苗的明确偏倚。

许多学区看到了 NFIP 方案中的这些缺陷而采用了不同的方案。对照组必须选自处理组所来自的同一母体：即父母同意接种疫苗的儿童们。否则，家庭背景的影响将与疫苗的影响相混淆。下一个问题是如何分配孩子到处理组或对照组。似乎需要人事判断，以使对照组与处理组具有相同的有关变量——家庭收入，或者儿童的一般健康状况、性格以及社会习惯。

但是经验显示人事判断常常引起实质性偏倚。最好采用一种

细致设计的机会方法。对于脊髓灰质炎试验，这种方法等价于为每一个儿童扔一次硬币，以 50%对 50%的机会把儿童分配到处理组或对照组。这种方法既客观又公正。机会法则保证了，在足够多对象的情况下，处理组与对照组关于所有重要的变量方面都将非常相近，而不管这些变量被识别与否。当采用了一个公正的机会方法来指派对象到处理组或对照组时，实验称之谓随机化对照的[3]。

脊髓灰质炎试验的另一个基本预防措施是使用安慰剂。给对照组的儿童注射盐溶液。在整个实验期间实验对象不知道他们是在处理组中还是在对照组中。因此他们的反应是对疫苗的，不是对处理的意见。实验对象光靠精神力量以抵御脊髓灰质炎似乎是不可能的。但是，医院中那些忍受手术后剧痛的病人已经在服用一种完全由中性物质制造的“止痛片”：大约 1/3 的病人经历过这种迅速减痛。[4]

还有一个预防：诊断医生必须确定在实验期间孩子是否感染了脊髓灰质炎。脊髓灰质炎的许多形式很难诊断，在不易确定的情况时医生的诊断可能受到事先知道孩子是否接种过疫苗的影响。因此，不告诉医生孩子是属于哪一组的。这就是双盲性：实验对象不知道他们是接受了处理还是注射了安慰剂，而且那些评估反应的人也不知道。这部分的脊髓灰质炎试验是随机化对照性的双盲实验。它大概是现有的最佳方案。

结果怎样呢？表 1 列出了在随机化对照实验中处理组与对照组的脊髓灰质炎病例比率(每十万人患病比率)。处理组的比例低得多，明显地证实了脊髓灰质炎疫苗的效果。

表 1 还显示 NFIP 对疫苗的研究有偏向于不利疫苗的方面。在随机对照实验中疫苗使脊髓灰质炎比率从十万分之七十一降低到二十八；NFIP 研究中表面上的缩减——从十万分之五十四降到二十五——太低了一点。偏性的主要来源是混淆。NFIP 处理组仅仅包含那些父母亲同意接种疫苗的儿童。而 NFIP 对照组还包含了那些父母亲不同意的儿童。对照组与处理组是不可比较的。

表 1 1954 年脊髓灰质炎疫苗试验结果。组的大小和各组中每十万人的脊髓灰质炎病例比率数字已进为整数。

随机对照双盲实验			NFIP 研究		
	人数	比率		人数	比率
处理	200 000	28	2 年级(接种疫苗)	225 000	25
对照	200 000	71	1、3 年级(对照)	725 000	54
不同意	350 000	46	2 年级(不同意)	125 000	44

来源：Thomas Francis，Jr. ，American Journal of Public Health Vol. 45(1955) PP. 1-63

随机对照双盲方案将偏性缩减到最小，这就是为什么只要有可能就使用它的主要理由。此外，这种方案有一个重要的技术性优点。为了查看根由，让我们扮演一会儿专门抬扛者，假定脊髓灰质炎疫苗毫无效果。那么，处理组与对照组的脊髓灰质炎比率之间的差异只能归因于机会。那有多大可能？

采用 NFIP 方案，其结果受到许多(从调研者观点)随机因素的影响：如哪些家庭自愿，哪些儿童在 2 年级，等等。然而，调研者缺乏足够的资料以估计这些结局的可能性。因此他们无法估计拒绝接受脊髓灰质炎比率间的差应归于这些偶然因素的可能性。另一方面，采用随机对照实验，当作出指派到处理组或对照组的分配时，机会以有计划的和简单的方式加入。

为清楚地说明此点，“专门抬扛者”的假设称疫苗没有效果。按照这个假设，少数几个儿童命中注定会感染上脊髓灰质炎；至于这几个患者是属于处理组还是对照组毫无干系。每个儿童一样依赖于抛一枚硬币以 50 对 50 的机会被安排进处理组或对照组。因此每 一个脊髓灰质炎病例有 50 对 50 的机会发生于处理组或对照组。

由此，两组中脊髓灰质炎病例数应当大致相等；任何差异应归于抛硬币中的机会变异。统计学家知道这类变异的变化。他们可以估计出由机会变异造成的差不同于观察到的差的可能性。并将在第 27 章中给出计算结果，这种可能性是极巨大的——十亿比一。

2. 静脉吻合分流术

在一些肝硬化病例中，病人可能开始出血且直到死亡。一种称之谓静脉吻合分流术的治疗方法是运用外科手术使血流改变方向。制造分流的手术既花时间又危险。这样的外科方法其得益是否值得冒此风险？为了评估这种外科治疗法的效果，曾经做过50多次研究。[5]其结果总结如表2。

表2 关于静脉吻合分流术的51次研究。仔细设计的研究显示外科手术价值很小或几乎没有价值。设计较差的研究夸大了外科手术的价值。

	热衷程度		
设计	显著	中等	无
无对照	24	7	1
对照，但不随机	10	3	2
随机对照	0	1	3

来源：Grace, Muench, and Chalmers, Journal of Gastroenterology Vol. 50 (1966) PP. 646－91

32次研究是无对照的：其中75%明显地热衷于分流术，得出结论为效益明确地超过风险。在15次有对照而在处理组或对照组的安排上并不随机化的研究中，只有67%明显地热衷于分流术。然而随机对照的4次研究证明分流外科几乎没有什么价值或者无价值。设计差的研究夸大了这种冒风险的外科手术的价值。

一种解释是：在一个无对照的实验中，或者根据病人的临床诊断来决定病人是接受处理或对照的实验中，存在一种自然的趋向：即仅仅治疗那些处于状态相对较好的病人。这使研究偏于有利处理。表2中所有三种类型研究，接受处理的病人中大约有60%在手术后3年仍然活着。在随机对照实验中，对照组中实验后活3年的百分比也大约为60%。但是在非随机实验的对照组中活3年的百分比只有45%。这显示了非随机实验中的偏性：健康状况较差的病人作为对照；而健康状况较好的人则选入处理组。对照组与处

理组是不可比较的。

3. 历史对照组

随机对照实验是很难做好的。一般，医生们常常采用并不比它好的其它方案。例如，他们对一组病人试验一种新治疗方法并将这些病人与“历史对照组”——即过去用老方法治疗的病人——相比较。问题在于处理组与历史对照组除了治疗方法之外可能在某些重要方面有所不同。静脉吻合分流术的“蹩脚对照”试验包括了一些带有历史对照的试验。(其它部分均为同时期对照，但对照组的指派不是随机的。)如第 2 节所述，设计很有关系。本节继续这方面的叙述。

例如，冠状旁道外科是对冠状动脉病广泛使用——且费用非常昂贵——的一种手术。Chalmers 与他的助手们鉴定了 29 例这类外科手术(表 3 第 1 行)。有 8 个随机对照试验，其中 7 个对该手术的价值是相当否定的。就比较来说，有 21 个使用历史对照的试验，其中 16 例持肯定态度。设计糟糕的研究比较热衷于手术的价值。表中其它行可以用同样方式理解并导致类似的结论。

表 3　关于一些研究事例的总结。利用随机对照试验和历史对照试验对 5 种疗法进行评估。试验的结果概括为对疗法的作用为肯定(＋)或否定(－)。

疗　法	随机对照		历史对照	
	＋	－	＋	－
冠状旁道外科	1	7	16	5
抗凝剂	1	9	5	1
5—FU	0	5	2	0
BCG(卡介苗)	2	2	4	0
DES(已烯雌粉)	0	3	5	0

注：抗凝剂用于治疗心脏病突发；5—FU 用于结肠癌化疗；卡介苗用于黑瘤；已烯雌粉用于防止流产。

来源：Sacks, Chalmers, and Smith. American Journal of Medicine Vol. 72(1982) PP. 233—40

关于冠状旁道外科的 6 例随机对照实验和 9 例具历史对照的

研究中，手术组与对照组的 3 年存活率如表 4 中所列。在随机对照实验中，手术组与对照组的存活率非常接近。这就是调研者对手术并不热衷的理由——它不能拯救生命。

表 4 在冠状旁道外科试验中手术病人与对照病人的 3 年存活率。随机对照实验不同于具历史对照的试验。

	随机对照	历史对照
手术组	87.6%	90.9%
对照组	83.2%	71.1%

注：有包括 9 290 个病人的 6 次随机对照实验；和包括 18 861 个病人的具历史对照的 9 次研究。

来源：见表 3。

现在，让我们来看使用历史对照的研究。手术组的存活率与前面所述随机对照实验的差不多相同。但是，对照组却有低得多的存活率。原因是他们一开始并不象那些被选去进行手术的病人那么健康。使用历史对照的试验偏向于赞同外科手术。随机化试验则避免了这类偏性。



现在我们回到表 3。最后一行值得更多的讨论。DES(已烯雌粉)是一种人工激素，用于防止自发性流产。Chalmers 与助手们找

到了 8 次试验来评估 DES。3 次为随机对照的，全部持否定意见：这种药无济于事。有 5 次是使用历史对照的研究，全部持肯定意见。这些较糟糕设计的研究偏于赞成该疗法。

医生们过去很少注意随机对照实验。即使在六十年代末期，他们每年给 50 000 名妇女分发这种药。正如后来的研究所揭示的那样，这是一个医学悲剧。如果在怀孕期间给母亲服药，DES 能在 20 年后带来灾难性副作用，可能引发她的女儿得一种极罕见的癌症（阴道透明细胞腺癌）。DES 已在 1971 年禁止给怀孕妇女服用。[6]

4. 小结

1. 统计学家经常使用比较法。他们想知道一种处理（如接种脊髓灰质炎疫苗）对一种反应（如患脊髓灰质炎）的效应。为了寻找出来，他们把一个处理组中的反应与一个对照组中的反应做比较。通常，不与其他事物比较就很难正确判断处理的效果。

2. 如果除了处理这一点之外，处理组与对照组完全相同，那么这两个组反应的差别就很可能应归因于处理的效果。

3. 然而，如果处理组相对于其他因素不同于对照组，则这些因素的影响就与处理的效果相混淆。

4. 为了保证处理组与对照组相同，调研者随机地将实验对象分到处理组或对照组。在随机对照实验中就是这样做的。

5. 只要有可能，给对照组一种性质中性但看上去很象处理的安慰剂。反应应当是针对处理本身而不是处理的想法。

6. 在双盲实验中，实验对象不知道自己在处理组还是在对照组；那些评估反应的人也不知道这点。这样就防止了反应中或评估中的偏性。

2
观察研究

你在那里所经历的不是一个实验，而是一种经验。

——R. A. Fisher 爵士(英国 1890—1962)

1. 引言

对照实验不同于观察研究。在一个对照实验中，是研究者决定谁将在处理组和谁将在对照组。与此形成对照，在观察研究中，正是实验对象安排他们自己到不同的组去：而研究者只是观察所发生的情况。

这句行话有点难懂，因为“对照”(Control)这个词有两种含意：

• 一个对照是一个没有接受处理的实验对象；

• 一个对照实验是一种由研究者决定谁在处理组中和谁不在处理组中的研究。

譬如，关于吸烟影响的研究必定是观察性的：没有人会仅仅为了讨好一位统计学家而去抽十年烟。但是，我们还是使用了处理—对照的思想。研究者比较吸烟者(处理或“暴露”组)与不吸烟者(对

照组）以确定吸烟的影响。

比较下来吸烟者结果很糟糕。在吸烟者中心脏病、肺癌和许多其他疾病比起不吸烟者来更为普遍。因此，在吸烟与疾病之间存在着一种强关联。但是这种关联并不等于因果关系。可能存在着某些隐蔽的混杂因素，它诱使人们吸烟——同时也使他们得病。倘若这样的话，中止吸烟就没有意义：因为它不能改变那些隐蔽的因素。

Joseph Berkson 与 R. A. Fisher 爵士这些统计学家们并不相信禁烟的证据，并提出一些可能的混杂因素。流行病专家（包括英国的 Richard Doll 爵士和美国的 E. C. Hammond，D. Horn，H. A. Kahn）进行了细致的观察研究证明了这些替代解释似乎是不可信的。总的说来，这些研究得出一有力的例证，那就是吸烟会引起心脏病、肺癌及其他疾病。如果你停止吸烟，你会活得更长。[1]

正如那个吸烟例子所说明的，观察研究是一种强有力的工具。但是它们也可能被严重误导。为了弄清楚混杂因素是否确成问题，找出如何选择对照也许将有所帮助。主要问题在于：对照组是否真正与处理组相象？如果存在混杂，必须对它进行某些工作——尽管不能指望做得尽善尽美。统计学家在观察研究中讨论对混杂因素的'控制'（Controlling）。这是词Control 的第 3 种用法。

控制的一种方法是对较小与较均匀的组作分开来的比较。例如，吸烟者与不吸烟者死亡率的粗略比较可能会使人误解，因为吸烟者大多为男性，而男人比起妇女来无论如何更易得心脏病。吸烟者与非吸烟者之间死亡率的差异也许应归因于性别的差异。为排除这种可能性，流行病专家对吸烟男性与不吸烟男性进行比较，对吸烟女性与不吸烟女性进行比较。

年龄是另外一个混杂变量：老年人有不同程度的吸烟习惯，并且他们患肺癌的风险较大。所以在吸烟者与不吸烟者之间的比较按年龄和按性别分开来进行：例如，将 55—59 岁的吸烟男性与 55—59 岁的不吸烟男性进行比较。这就控制了年龄与性别。好的

观察研究应控制混杂变量。

最后，大多数观察研究没有象吸烟问题那样成功。研究方案也许是专家设计的，但专家们也会出错。发现弱点与其说是科学不如说是艺术，且常常依赖于研究以外的信息。下面的各种例子将揭示其道理。

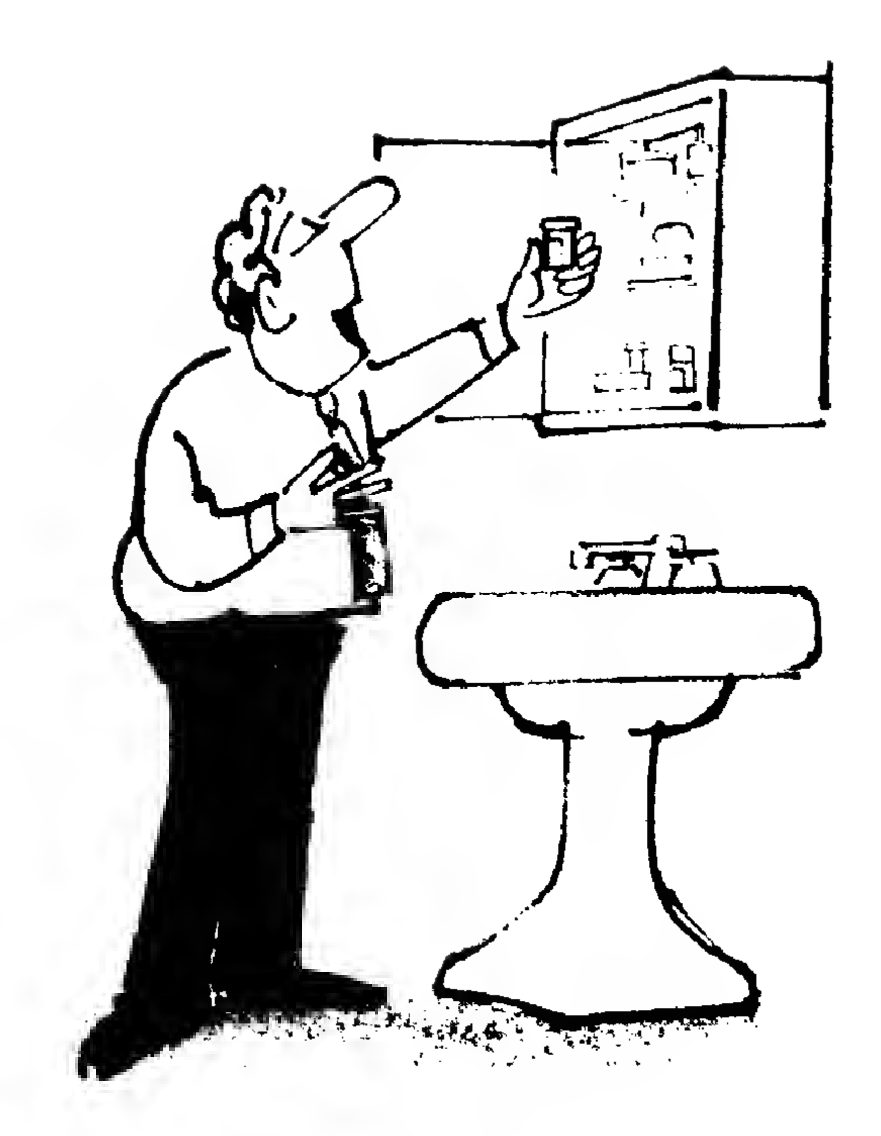
2. 安妥明试验

冠心病药物研究项目是一随机对照双盲实验，目的是评估 5 种防止心脏病突发的药品。实验对象是心脏有问题的中年男子。在 8 341 名实验对象中，随机地指派 5 552 名到药物组，2 789 名到对照组。药与安慰剂（乳糖）放入相同的胶囊中分别分发给药物组与对照组中的对象。这些病人受到 5 年的跟踪观察。

接受试验的药物之一是安妥明，其作用是降低血液中的胆固醇水平。不幸地，这种治疗没有拯救任何生命：在整个跟踪期间安妥明组中有大约 20%病人死亡，相比较对照组有 21%病人死亡。这次失败的一个可能的理由——安妥明组中许多对象不遵医嘱服药。

我们称服用 80%以上所配予的药（或安慰剂）的对象为该试验的“坚持者”。在安妥明组中，坚持者中间 5 年跟踪期内的死亡率仅为 15%，相比较不坚持者的死亡率为 25%（表 1）。这看来似乎是药有效性的有力证据。但是，小心谨慎是应该的，这一特殊的比较是观察性的，不是实验性的——尽管数据是在实验进行中收集得到的。毕竟，调研者并不决定谁将是试验的坚持者和谁不是坚持者；这是由对象自己所决定的。

兴许坚持者与不坚持者除了在他们所服的药量之外在其它方面有所不同。为找出这些方面，调研者比较了对照组中的坚持者与不坚持者。记住，这是一个双盲实验。因此对照者们不知道自己是在服用医生想配与的药物还是安慰剂。在安妥明组中的对象也是这样。在两个组中坚持性的心理基础是完全一样的。



“坚持或不坚持正是问题所在。”

表 1 安妥明试验。接受试验的人数，和 5 年跟踪期间的死亡百分比。坚持者服用了 80%以上所配予的药。

	安妥明		安慰剂	
	人数	死亡率	人数	死亡率
坚持者	708	15%	1 813	15%
不坚持者	357	25%	882	28%
整个组	1 103	20%	2 789	21%

注：关于坚持者的资料，安妥明组中少了 38 个病人，安慰剂组中少了 94 个病人。死亡率则来自所有病例。

来 源：The Coronary Drug Project Research Group, New England Journal of Medicine vol. 303(1980) pp. 1038—41。

在对照组中也是这样，坚持者的情况较好一些——他们中在 5 年期内只有 15%的人死亡，相比较不坚持者中有 28%的人死

亡。结论：

(i)安妥明没有效果。

(ii)坚持者与不坚持者有所不同。

也许，坚持者更关心他们的健康，在总的方面照顾他们自己较当心一些。那就解释了为什么他们服用胶囊以及为什么他们活得长些[②]。观察比较可能会使人误解。安妥明试验中的调研者们可能非常仔细，他们发现了比较坚持者与不坚持者中的错误。

3. 实例

例1."西班牙医生 Gaspar Casal 于十八世纪在欧洲首次观察到糙皮病，他发现这种病在 Asturias 非常贫穷的居民中是体弱多病、伤残、夭折的一个重要原因。在以后的若干年内，许多作者…描述了意大利北部，尤其是来自伦巴第平原的农民的同样状况。到了十九世纪初，糙皮病进展性地蔓延横贯欧洲，犹如一条腰带，引起法国南部，奥地利，罗马尼亚和土耳其帝国领地内数以千计的人们进展性发生体力恶化和智力衰退。欧洲之外，在埃及和南非也发现了糙皮病，到了二十世纪头十年，糙皮病在美国，尤其是在南部蔓延泛滥…"[③]

糙皮病侵袭一些乡村似乎甚于其他地方。即使在受侵袭的村庄里有许多家庭幸免于难，但是有些人年复一年地患糙皮病。患病家庭的环境卫生条件是原始的，到处有苍蝇。至少在欧洲，一种吸血蝇与糙皮病有着同样的地理范围；而吸血蝇在春天最为活跃，恰恰是发生糙皮病例最多的季节。

许多流行病专家断定这种疾病是传染性的，并且——象疟疾，黄热病，或班疹伤寒那样——由昆虫传染。这个结论正确吗?

讨论。大约1914年初，美国流行病专家 Joseph Goldberger 通过一系列观察研究和实验证实了糙皮病是由不良的饮食引起而不是传染性的。这种疾病可以通过食用富含一种 Goldberger 称之为 P—P(防糙皮病)因子的食物而预防或治愈。1940年以来，美国销

售的大部分面粉中在其他的维生素里添加了P—P因子。(P—P因子的商业称号为“烟酸”。)

烟酸天然地存在于肉,奶,蛋,一些蔬菜,和某些谷物之中。但是,相对地玉米几乎不含烟酸。在糙皮病地区,穷人们吃玉米——其它的食物并不多。一些村庄和一些家庭比其他人还要穷,饮食更为限制。那就是他们受到疾病较严重侵害的原因。苍蝇是贫穷的一种标志,而不是糙皮病的起因。相关联不等同于因果关系。

例2. 子宫颈癌与包皮切除。子宫颈癌多年来一直是妇女中最普遍的癌症之一。许多流行病专家一直致力于查明该病的病因。他们发现在一些不同的国家中,犹太人中很少见子宫颈癌。他们也发现在穆斯林中少见此病。在五十年代,一些研究者推断男性的包皮切除是一种防护性因素。他们正确吗?

讨论。除了切割包皮之外,在犹太人或穆斯林与其它团体人员之间还存在着差异。结论是子宫颈癌是一种性病,它可以通过接触而传播。诱发物被认为是一种病毒。比起其他人来,有些妇女性欲较旺且有较多情人。那似乎是使她们得子宫颈癌率高于他人的原因。早期的研究没有注意到这个混杂变量,从而得出错误的结论[④]。(癌症的发展须经历一个长时期;三十或四十年代性行为就是原因。)

例3. 超声波与低初生婴儿重量。近年来可利用超声波检测子宫中的婴儿。关于实验室动物的一些实验证实了超声波检测可能引起低初生重量。这个结论倘若对人类也是真的话,就有被关心的理由。研究工作者曾在巴尔的摩的Johns Hopkins医院进行观察研究以查明这一点。

当然,暴露于超声波下的婴儿除了暴露这一点以外在许多其他方面与不暴露于超声波下的婴儿有所不同,这是一件观察研究。研究工作者们发现了若干混杂变量并对它们作出调整。纵然如此,还存在一种关联关系。平均来说,暴露于超声波下的胎儿比起不暴露的胎儿有较低初生重量。这是超声波引起低初生重量的根据吗?

讨论。产科医生在情况不正常时建议超声波检查。研究工作者归纳为超声波检查与低初生重量具有共同的原因——成问题的怀孕。后来，做了一个随机对照实验得到了更确凿的证据。如果有什么差别的话，那就是超声波检查是防护性的[5]。

例4. 救济会与自杀。在1964－1970年这个时期，英国的自杀率降低了大约1/3。在这个时期内，一个称为“救济会”的志愿性福利组织迅速地扩展起来。一位研究者认为救济会与自杀的减少有关系。他做了一次观察研究以证明之。研究依据15对城镇进行；为了控制混杂变量，每对中的城镇在认为是重要的变量方面是互相匹配的。每对中有一个城镇有一救济会的分支而另一个城镇则没有。大体上，有“救济会”的城镇自杀率较低。因此“救济会”阻止了自杀。难道不是这样吗？

讨论。第2位研究者利用较大的样本和更仔细的匹配重复了这个研究；没有发现什么两样。此外，在二十世纪七十年代(第1位研究者发表他的论文之后)尽管“救济会”继续扩充，但是自杀率是稳定的。六十年代自杀率的下降可以通过为取暖及烹调所用的煤气转换为天然气这件事所作出较好的解释。天然气毒性小得很多。事实上，六十年代早期的自杀大约有1/3是利用煤气。六十年代末，几乎没有这类事例，从而解释了自杀减少的原因。向天然气的转换完成了，因此利用煤气的自杀率不可能进一步减少较多。最终，直到六十年代过完——尽管有“救济会”——利用非煤气方法的自杀率接近于常数。“救济会”是个好组织，但在自杀率方面他们似乎没有多大影响。不管多么仔细地去实施，观察研究不是实验[6]。

4. 研究生入学的性别偏倚

简单地评估一下，观察研究中一个烦恼的来源是观察对象中除了处理之外在许多重要方面有所不同。有时候这些差异可以通过对较小和较均匀的子组进行比较而加以调整。统计学家称这种

技巧为“控制混杂因素”——Control 一词的第 3 种含意。

例如，加利福尼亚大学伯克莱分校的研究生部曾就入学的性别偏倚作了一次观察研究[⑦]。在此项研究期间，有 8 442 名男生和 4 321 名女生申请报考研究生。录取了大约 44%的男生和 35%的女生。采用取百分数的办法调整了在男女申请者数目上的差异：每 100 名男生中有 44 名录取而每 100 名女生中录取 35 名。

假定大体上男女有同样好的资格(没有相反的迹象)，那么录取率方面的差异看来好象是说明在入学程序中男女遭受不同待遇的一个强有力的依据。大学似乎以 44 对 35 的比例偏爱男生。

每个专业做自己的研究生录取工作。分别地观察它们，应能辨认出那些岐视妇女的专业。在这一点上出现了一个困惑。一个专业一个专业地观察，似乎不存在任何岐视妇女的偏性。有些专业喜欢男生，而其他专业喜欢女生。总的来说，如果存在偏性的话，它倒是不利于男生的。会发生什么事情呢?

整个研究生部包含了 100 多个专业，数目太多而无法在这里一一考虑。但是，6 个最大的专业一起共计占全校总申请数的 1/3 以上，以这些专业为范本于整个学校是典型的。表 2 罗列了这些专业中的每一个专业的男女申请人数以及录取的百分数。

表 2 加利福尼亚大学伯克莱分校 6 个最大专业的研究生入学资料。

	男		女	
专业	申请人数	录取百分数	申请人数	录取百分数
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

注：大学规定不允许这些专业暴露名称。

来源：加利福尼亚大学伯克莱分校研究生部。

在每个专业，录取的女申请者的百分数大约等于男申请者录取的百分数。唯一的例外是专业A，那里呈现了岐视男生现象：它录取了申请女性的82%而仅录取申请男性的62%。看起来最偏于反对妇女的系是E。它录取了申请男性的28%和申请女性中的24%。差仅为4%。然而，当6个专业合在一起时，它们录取了男申请者中的44%，而女申请者中仅录取30%——差为14%。



"是的，从表面上似乎呈现性别偏性，
但是让我们提出如下问题…"

这的确是自相矛盾的，但这里给出解释：

- 头两个专业容易考取。50%以上的男生申请这两个专业。
- 其他4个专业较难考取。90%以上的女生申请这4个专业。

因此男生申请容易的专业而女生则申请较难的。存在着由专业选择所产生的影响，它与由性别引起的影响混杂在一起。当专业的选择受控制，如表2中那样，男女的入学率之间没有多大差别。

在许多观察研究中，有可能通过一个类似的过程控制混杂因素：对结构相似的子组作单独的比较。

表 3 由表 2 而得的申请者总数。

专业	申请者总数
A	933
B	585
C	918
D	792
E	584
F	714
	4 526

技术性注。表 2 难于读懂，因为它比较了 12 个入学率。一个统计学家通过计算一个关于男生的综合入学率和另一个关于女生的综合入学率可以归纳表 2，不过在计算中应调整申请率中的性别差异。方法为分别对男生和女生取某种类型的平均入学率。普通的平均数将行不通，因为它忽略了各系之间在大小方面的差异。取而代之，将使用一种关于入学率的加权平均数，权数为各系（男和女）申请者的总数；见表 3。

男生的加权平均入学率是

$$\frac{0.62\times933+0.63\times585+0.37\times918+0.33\times792+0.28\times584+0.06\times714}{4\ 526}$$

得到 39%。同样地，女生的加权平均入学率是

$$\frac{0.82\times933+0.68\times585+0.34\times918+0.35\times792+0.24\times584+0.07\times714}{4\ 526}$$

得到 43%。在这些公式中，男女的权数是一样的：它们为来自表 3 的总数。男女的入学率是不同的：它们是来自表 2 的比率。最终比较：男生的加权平均入学率是 39%，而女生的加权平均入学率是 43%。加权平均数控制了混杂因素——专业的选择。这些平均数使人想到如果有什么问题的话，那是录取过程不利于男生。

习题 A

1. 在美国，1985 年所有的病例中有 210 万人死亡，对比 1960 年的 170 万——几乎增加了 25%[8]。公众的健康状况在这段时期内变糟糕了吗？回答是或否，并解释之。

2. 从第 1 章中的表 1，NFIP 对照组比随机实验中的对照组有较低的小儿麻痹症（脊髓灰质炎）发病率。为什么？

3. NFIP 研究是在盲式情况下进行的吗？那将会怎样影响结果？简短地讨论。

4. 脊髓灰质炎疫苗现场试验仅仅在一定的实验区域（某些学区）内进行，它们是由公共卫生总署与地方当局磋商而选择的[9]。在这些区域内，有大约 300 万 1，2，或 3 年级儿童；而在美国有大约 1100 万儿童在那几个年级。在该实验区域内，脊髓灰质炎的发病率大约是 25%，高于这个国家的其他地方。脊髓灰质炎疫苗现场试验引起了儿童得脊髓灰质炎而不是预防吗？回答是或否，并作出简短的解释。

5. Linus Pauling 认为维生素 C 预防感冒，并且也可以治愈感冒。Thomas Chalmers 和他的助手们做了一个随机对照双盲实验进行核实[10]。实验对象是国家卫生研究所的 311 名志愿者。这些对象被随机地指派到 4 个组中的一个：

组	预　防	治　疗
1	安慰剂	安慰剂
2	维生素 C	安慰剂
3	安慰剂	维生素 C
4	维生素 C	维生素 C

所有对象一天发给 6 颗胶囊用于预防，如果他们倒霉患了感冒则一天还发另外 6 颗胶囊用于治疗。但是，在第 1 组中的两组胶囊只含安慰剂（乳糖）。在第 2 组中，用于预防的胶囊含维生素 C 而治疗用的胶囊则充进了安慰剂。第 3 组与第 2 组情况相反。而在第 4 组中，所有的胶囊都注满了维生素 C。

在试验期间有相当高的退出率。这个比率在头 3 组中明显地高于第 4 组。调研者们注意到了这个事实并发现了原因。原来是许多对象违反了“盲式”规则。（那是相当容易做的；你直接打开一颗胶囊并尝一下里面的东西；维生素 C——抗坏血酸——是酸的。）那些得安慰剂的对象更可能退出。

调研人员分析了那些仍处于“盲式”状况的对象的数据，发现维生素 C 没有效果。可是对于那些违反“盲式”规则的人，分在第 2 及第 4 组的这种

人得感冒者最少；分在第 3 及第 4 组的这些人倘得了感冒，经受的期间最短。你怎样解释这些结果？

6.（假设的）在冠心病药物研究项目（第 2 节）中的其它药物之一是烟酸[11]。假定有关烟酸的结果报告如下。其中有些结果看来是错的。是哪些，为什么？

	烟酸		安慰剂	
	人数	死亡率	人数	死亡率
坚持者	558	15%	1 813	15%
不坚持者	487	27%	882	28%
整个组	1 096	22%	2 789	21%

注：关于坚持者的资料，烟酸组中少了 51 人，安慰剂组中少了 94 个。

7.（假设的）临床试验中，数据收集通常从“基线”开始，即当试验对象被接纳试验中来但尚未被指派到处理组或对照组去之前的时候。数据收集工作一直持续到跟踪结束。两个关于预防心脏病突发的临床试验报告了有关吸烟的基线数据，如下所示。这些试验中的某一个的随机化没弄好。是哪一个，为什么？

		人数	吸烟百分比
(i)	处理	1 012	49.3%
	对照	997	69.0%
(ii)	处理	995	59.3%
	对照	1 017	59.0%

8. 乳腺癌是美国妇女中最常见的恶性肿瘤之一。如果很早地发现——在癌扩散之前——则治疗成功的机会更好。能否说检查计划明显加快了发现吗？第一次大规模试验由全纽约健康保险规划于 1963 年开始实施。实验对象（规划中的所有成员）为 62 000 名 40 至 64 岁的妇女。这些妇女随机地被分为两个大小一样的组。在处理组中，妇女们受到鼓励去接受年度检查，包括由医生和 X-光进行的检查。处理组中大约 20 200 名妇女接受了检查；但是有 10 800 名妇女拒绝检查。对照组的人被提供普通健康检查。所有妇女被跟踪若干年。头 5 年的结果如下表所示[12]。

HIP 检查试验头五年里的死亡人数，附死因。比率是每 1 000 名妇女的。

		死因			
		乳腺癌		其他	
		人数	比率	人数	比率
处理组					
检查	20 200	23	1.1	428	21
拒绝	10 800	16	1.5	409	38
总数	31 000	39	1.3	837	27
对照组	31 000	63	2.0	879	28

(a)检查拯救了生命吗？表中哪几个数字证明你的观点？

(b)在拒绝检查的妇女中的死亡率(来自所有原因)是接受检查妇女的死亡率的大约两倍。检查使死亡率降低了一半吗？请简短地解释。

(c)该研究是“盲式”下进行的吗？

9. (续习题 8)在 HIP 试验的第 1 年中，在“检查”组中发现了 67 例乳腺癌；“拒绝”组中发现了 12 例；而在对照组中发现了 58 例。对下述结论回答正确还是错误，并简短地解释：检查引起乳房癌。

10. 根据许多观察研究，子宫颈癌在那些接触过疱疹病毒的妇女中间是较常见的[13]。是否可直接地推断这种病毒引起了子宫颈癌？

11. 一项观察研究发现那些有规律地进行体育锻炼的妇女比起其他妇女较少发生自发性流产[14]。体育锻炼降低了自发性流产的风险吗？

12. 在十八世纪，通过给病人放血治疗黄热病。当时一位著名的医生 Benjamin Rush 博士写道：

    我从每一次抽少量的血开始。血的显露以及放血对身体的效应使我相信放血治疗的安全性和有效性。在这之前我从未经历过象现在那种在期待我的治疗成功中所感受的异常的兴奋…。当我加上一小段摘自 1793 年 9 月 10 日我的笔记中的摘录时读者将不会感到惊讶：“感谢上帝，在我出诊或给予配方的 100 个病人中，至今无人不幸。”[15]

    在 Rush 的研究中，存在的某些设计问题是什么？

13. 一所虚构的大学有 A 和 B 两个系。有 2 000 名男申请者，他们中的各一半申请各个系。有 1100 名女申请者：100 名申请 A 系而 1 000 名申请 B 系。A 系录取了申请该系的男生的 60%和女生的 60%。B 系录取了申请该系的男生的 30%和女生的 30%。“对每个系来说，男生录取的百分数等于女

生录取的百分数；两个系合在一起必定也是这样。”正确还是错误，并简短地解释。

习题 14 和 15 是为下一章作准备而设置。在练习中请勿使用计算器。只要记住“百分之…”意指“每一百中”。例如，398 人中的 41 人大约为 10%。理由：398 中的 41 接近于 400 中的 40，即 100 中的 10，那就是 10%。

14. 指出下列中的每一个大约是 1%，10%，25% 还是 50%——

(a)398 中的 39　　(b)407 中的 99

(c)209 中的 57　　(d)197 中的 99

15. 在 1987 年秋季一所大学中学初等统计学的学生中，446 名学生中的 46 名报告了其家庭收入为一年 10 000 美元至 20 000 美元。

(a)大约有多少百分数的学生其家庭收入为一年 10 000 美元至 20 000 美元？

(b)猜测其家庭年收入在 15 000 美元至 16 000 美元范围内的学生的百分数。

(c)猜测其家庭年收入在 16 000 美元至 17 000 美元范围内的学生的百分数。

(d)猜测其家庭年收入在 17 000 美元至 19 000 美元范围内的学生的百分数。

这些习题的答案在第 670—672 页上。

5. 复习题

复习题可能包含前面几章的内容。

1. 在美国，1985 年有 19 893 人遭到凶杀，与 1970 年的 16 848 人遭到凶杀相比——几乎增加了 20%[16]。“这些数字揭示了在 1970—1985 年期间美国变成为一个更多暴力的社会。”正确还是错误，并简短地解释。

2. 国家公路与交通安全管理署分析了 1987 年新车盗窃，并同该年的新车生产数量相比较[17]。

(a)在豪华车方面，生产的 157 374 辆 Cadillac Deville 中的 970 辆被盗；而生产的 68 106 辆 Chrysler New Yorker 中有 747 辆被盗。正确还是错误，并作出解释：由于 970 大于 747，因

此小偷喜欢 Cadillacs。

(b)在 Chevrolet 装配线,生产的 128 056 辆 Camaros 中有 3 333 辆被偷。因此 Camaros 的被盗率是 128 056 中的 3 333,或者是 2.6%。生产的 74 739 辆 Monte Carlos 中有 1 516 辆被盗,因此被盗率是 2.0%。正确还是错误,并作出解释:Camaros 的失窃率较高些,因为它们有那么多的产品。

3. 从第 1 章中的表 1,那些父母拒绝参加随机对照脊髓灰质炎试验的孩子以每 100 000 人中有 46 人的比率得脊髓灰质炎。另一方面,那些父母同意参加试验的孩子,处理组与对照组合在一起,以每 100 000 人中有 49 人的较高比率得脊髓灰质炎。假如该现场试验在下一年重复进行。在这些数字的基础上,某些家长拒绝允许他们的孩子参加试验而去面临脊髓灰质炎的较高风险。他们对吗? 回答是或否,并简短地解释。

4. 在美国,死亡的主导原因之一是心脏的冠状动脉病,这种病表现为通向心脏的主要动脉衰竭。正如在第 1 章的第 3 节中所讨论,这种病可以使用冠状旁道外科手术来治疗。在第一批手术试验中的某一次,Daniel Ullyot 博士和他的同事们对病人的一个试验组实施了冠状旁道外科手术;98%的手术病人又活了 3 年或 3 年以上。以往的研究表明那些得到常规处理的病人只有 68% 活了 3 年或 3 年以上。(常规处理使用药物和特殊的忌口饮食来减小血压和消除沉淀在动脉中的脂肪。)一篇报刊文章把 Ullyot 的结果说成是"惊人的"[18]。简短地评议此事实。

5. 在一确定的地区(苏格兰的 Lanarkshire)进行了一次实验来确定对学童提供免费牛奶的效果。[19]在每所学校选择一些儿童作为处理组并给予免费牛奶;其他儿童被选作对照而得不到牛奶。关于指派去处理组或对照组的工作是随机地进行的,使得这两个组在家庭背景与健康方面是可比较的。但是,单靠偶然性在这两组中仍然会有小的差异,这一点令人担忧。因此出于使这两个组相等同这一目的,允许教师运用自己的判断转换处理组与对

照组之间的儿童。让教师运用自己的判断这种方式是明智之举吗？回答是或否，并作简短的解释。

6. 公共卫生总署以典型家庭的一组大样本研究吸烟关于健康的影响[20]。对于每一年龄组中的男性和对于每一年龄组中的女性，那些从未吸烟的人平均来说比现时吸烟者稍微健康一些，而现时吸烟者比起以前吸烟者平均来说要健康得多。

(a)为什么他们分别地研究男性和女性以及不同的年龄组？

(b)教训似乎是你不应该开始吸烟，但是一旦你吸烟了，不要戒掉。简短地评议这个结论。

7. 有一种少见的使得食物尝起来是腐败味的神经病（原发性厌物症）。有时候使用硫酸锌来治疗。一组研究人员做了两次随机化对照实验来检验这种治疗法。在第一次试验中，病人不知道他们得到的是硫酸锌还是安慰剂。但是评价结果的医生知道。在这次试验中，分在硫酸锌组的病人的病况显著地得到改善；安慰剂组则几乎没有改善。第二次试验是在双盲情况下进行：病人与做评估工作的医生均不被告知病人得到的是药还是安慰剂。在第二次试验中，硫酸锌没有什么效果。应该配予硫酸锌去治疗这种疾病吗？回答是或否，并简短地解释。

8. （续前面习题）。第二次试验采用了称之为“交叉”的方案。对象们随机地被指派到如下四个组中的一个：

安慰剂	安慰剂
安慰剂	硫酸锌
硫酸锌	安慰剂
硫酸锌	硫酸锌

在第一组中，整个实验过程试验对象一直使用安慰剂。在第二组，病人们开始时使用安慰剂；但实验进行了半程后，他们转换为使用硫酸锌。类似地，在第三组中，病人开始时使用硫酸锌而中途转为安慰剂。在最后一组中，他们始终使用硫酸锌。病人们知道研究方案，但未被告知自己被指派到哪一组。在实验的前半

过程中有些病人的状况没有得到改善。平均来说，所有四个组的病人在实验的后半过程里均显示出某种程度的改善。这种现象如何给予解释？

9. 关于静脉吻合分流术（第 1 章第 2 节），在蹩脚设计的试验中的对照者比随机对照实验中的对照者其幸存可能性较差。在一个蹩脚设计的研究中作为对照者是危险的吗？回答是或否并予解释。如果你的答案为否，说明在幸存率方面的差异的原因是什么？

10. 根据在加利福尼亚州，Walnut Creek 的 Kaiser Permanente 所做的观察研究，即使在对年龄，教育，婚姻状况，宗教，和吸烟作出调整之后，口服避孕药的使用者比不使用者有较高比例的子宫颈癌。研究人员得出结论为（女用）口服避孕药引起了子宫颈癌[22]。他们这样做对吗？回答是或否，并简短地解释。

11. 许多研究揭示了在少年犯罪与家庭大小之间存在着一种强关联：来自大家庭的儿童比起来自小家庭的儿童更可能成为少年犯。既使在种族，宗教，和家庭收入这些因素得到控制的情况这种关系依然存在，使人联想到家庭大小是构成少年犯罪的一个起作用的因素。

    一项研究发现，通过与全体居民的比较，高百分数的少年犯罪发生在排行居中的儿童——就是说，既不是第一个出生的小孩，也不是最末出生的小孩。既使在种族，宗教，和家庭收入这些因素得到控制的情况，这种关系依然存在。因此，作为一个排行居中的小孩，似乎是构成少年犯罪的一个起作用的因素[23]。难道不是吗？回答是或否，并予解释。

12. （假设的）在某城进行一项研究以确定在选举态度上党派成员的作用。整个城市划分成选区。在每个选区内，参加投票的注册民主党人的百分数高于参加投票的注册共和党人的百分数。正确还是错误，并予解释：该城作为一个整体，参加投票的注册民主党人的百分数必定高于参加投票的注册共和党人的

百分数。

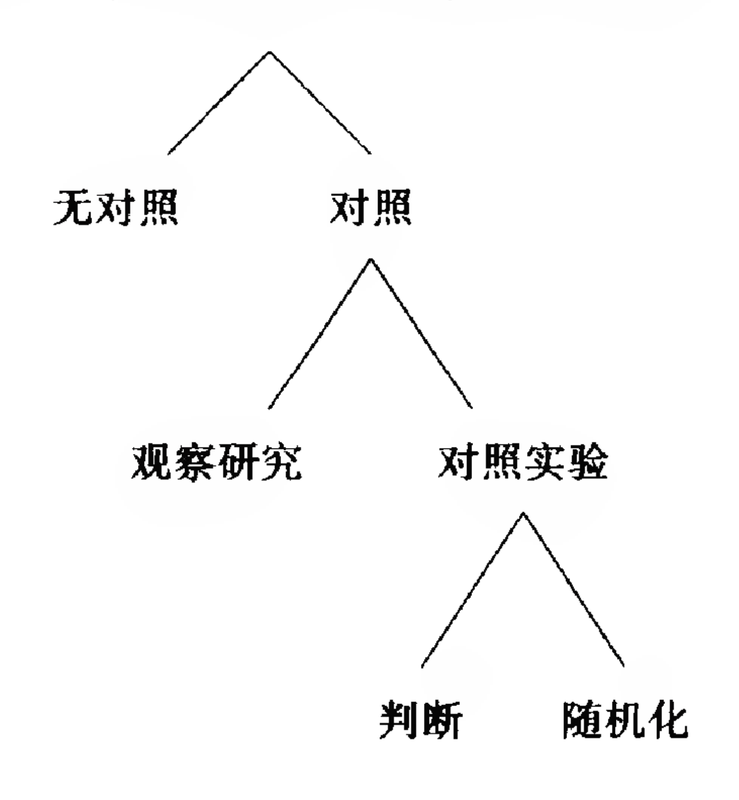
6. 小结

1. 在观察研究中，调研人员不指派受试验对象去处理组或对照组。接受试验的某些对象拥有其效果正在被研究的条件，这就是处理组。其余的形成对照组。例如，在有关吸烟的研究中，吸烟者形成处理组而不吸烟者则为对照物。

2. 观察研究可以建立关联性：一事与另一事相关联。但是关联性不一定是因果关系。

3. 在一项观察研究中，处理的效果可能与某些因素的效果混杂在一起，那些因素在一开始就影响到受试验对象进入处理组还是对照组。因此观察研究在有关因一果关系上可能或多或少地被引入歧途。

4. 当考察一项研究时，提出下面的问题。存在任何的对照组吗？若有，受试验对象如何被指派去处理组或对照组：通过一个在调研人员控制下的过程（对照实验），还是通过一个调研人员控制之外的过程（观察研究）？如果它是一项对照实验，指派工作是使用了机会技巧（随机对照）还是依赖于调研人员的判断？



5. 对观察研究和对非随机化对照实验，试着找出受试验对象是如何来到处理组或者对照组的。这两个组可比较吗？不同吗？什么因素与处理相混杂在一起？作什么样的调整使得混杂消除？它们合理吗？

6. 在一项观察研究中，混杂因素有时候可以通过比较关于因素相对类似的较小的组而加以控制。

第二部分　描述性统计

3
直方图

成人喜欢数字。当你告诉他们你已经交上一个新朋友时，他们决不会问你任何有关基本事宜的问题。他们不会对你说："他的声音象谁？最喜欢什么游戏？他收集蝴蝶吗？"取而代之，他们会问："他有多大年龄？有几个兄弟？体重有多少？他的父亲挣多少钱？"仅仅从这些数字，他们才认为已经了解了有关他的一切。

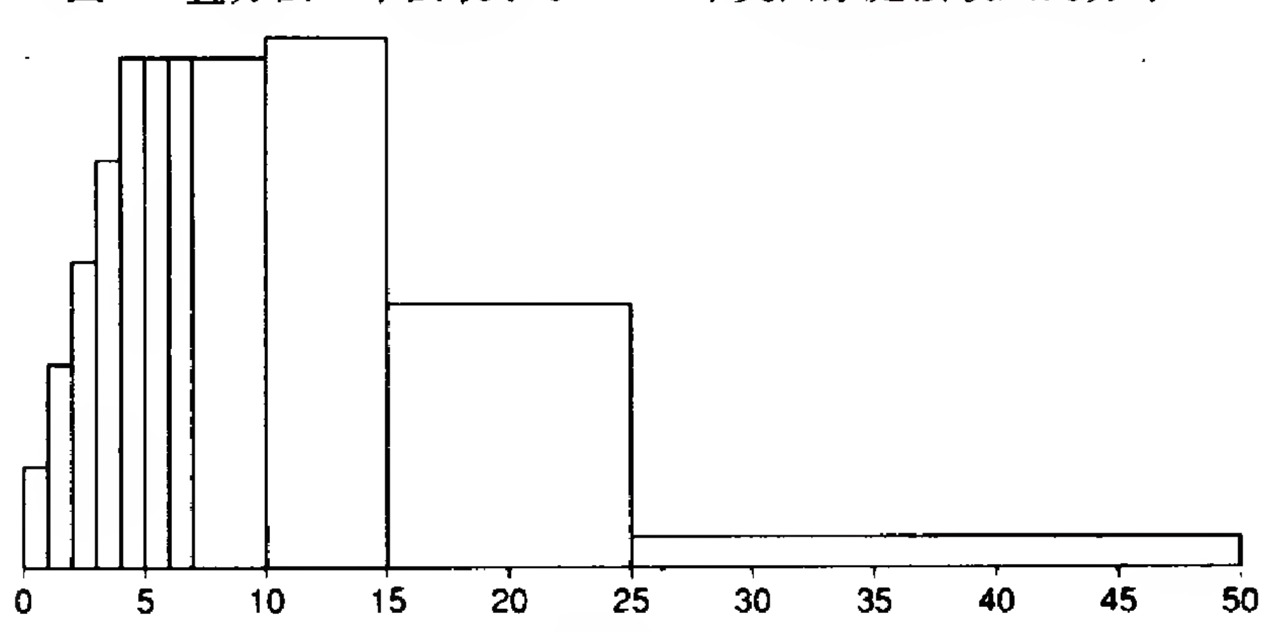
——The Little Prince①

1. 引言

在美国，收入的分布情况怎样？少数民族群体的情况有多差？政府统计通过现场人口调查提供了某些信息。该调查的机理将在第六部分讨论；每一个月，该调查提供了大约 50 000 个美国家庭的一个有代表性的剖面。在每年的三月，要求这些家庭报告在过去的一年内他们的收入。我们将观察 1973 年的结果。当然，这些数

据必须以某种方式加以概括。没有人想去看 50 000 个数字。为了概括数据，统计学家经常使用一种叫做直方图的图像。关于收入数据的直方图展示如图 1。

图 1 直方图。本图表示了 1973 年美国家庭按收入的分布。



来源：现场人口调查[②]。

本节讲解如何看懂直方图。首先，它与大多数其他图像不同，没有纵向刻度，直方图不需要纵向刻度。接下来考虑的事是水平刻度。它以千美元为单位表示收入。图像由一批块形构成。第一个块形的底边包含了从 0 至 1 000 美元的区域，第二个块形的底边从 1 000 美元到 2 000 美元，等等，直到最后一个块形，它包含了 25 000 美元至 50 000 美元的区域。这些区域称之为小组区间。图像被绘制得使每一块形的面积正比于其收入在所对应的小组区间内的家庭数。

直方图是用面积而不是用高度来表示数。

如此看来，直方图不同于条形图。

为了看清这些块形是如何起作用的，更细致地观察图 1。大约多少百分数的家庭挣得收入在 10 000 美元与 15 000 美元之间呢？该区间上方的块形大约相当于所有块形总面积的 1/4。因此大约 1/4，或 25%的家庭具有那个范围内的收入。

举另外一个例子。有较多的家庭收入在 10 000 美元至 15 000

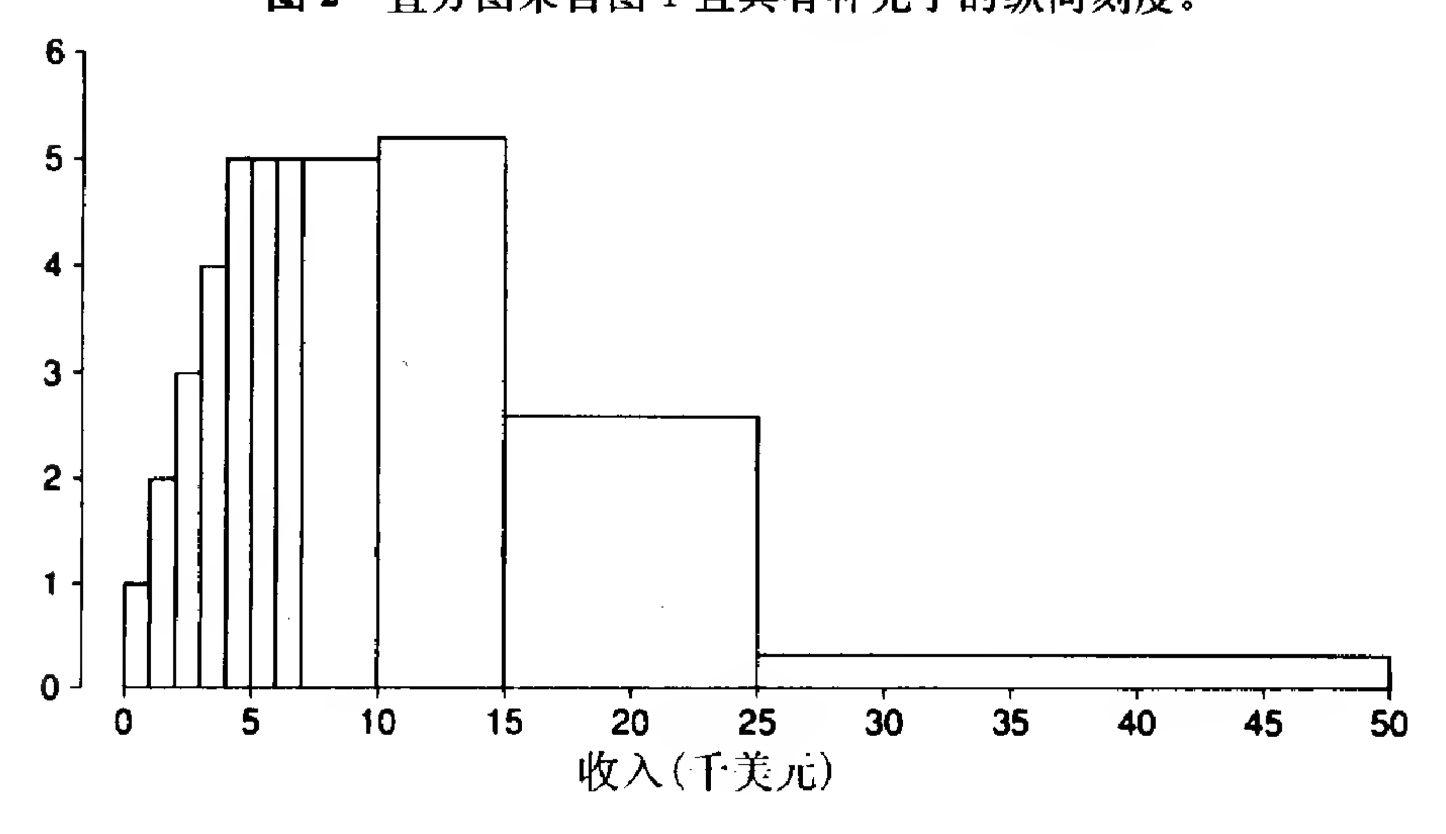
美元之间，还是较多的家庭收入在 15 000 美元至 25 000 美元之间?第一个区间上方的块形较高一些，而第二个区间上方的块形较宽一些。两个块形的面积大致相同，因此挣 10 000 至 15 000 美元的家庭数大约等于挣 15 000 至 25 000 美元的家庭数。

最后一个例子，取收入在 7 000 美元以下的家庭的百分数。它最接近于 10%，25%，还是 50%? 由目测，在 0 与 7 000 美元之间的直方图下的面积大约是总面积的 1/4，因此百分数最接近于 25%。

图 1 中的水平轴停止在 50 000 美元。所挣超过此数的家庭怎么样呢? 直方图简单地忽略了他们。当然，1973 年只有 1%的美国家庭具有超过那个水平的收入；大多数家庭在图中得到表示。

这方面，更多了解直方图的一个好途径是做一些习题。为了帮助你判断块形的大小，图 2 给出了如图 1 一样的直方图，但具有补充了的纵向刻度。

图 2 直方图来自图 1 且具有补充了的纵向刻度。



习题 A

1. 图 2 中大约 1%的家庭收入在 0 至 1 000 美元之间。估计收入——
   (a)在 1 000 与 2 000 美元之间

(b)在 2 000 与 3 000 美元之间

(c)在 3 000 与 4 000 美元之间

(d)在 4 000 与 5 000 美元之间

(e)在 4 000 与 7 000 美元之间

(f)在 7 000 与 10 000 美元之间的家庭所占的百分数。

2. (a)在图 2 中，有较多的家庭所挣在 6 000 与 7 000 美元之间呢，还是在 7 000 与 8 000 美元之间？或者它们的家庭数目大致相同吗？

(b)在图 2 中，有较多的家庭所挣在 10 000 与 11 000 美元之间呢，还是在 15 000 与 16 000 美元之间？或者它们的家庭数目大致相同吗？

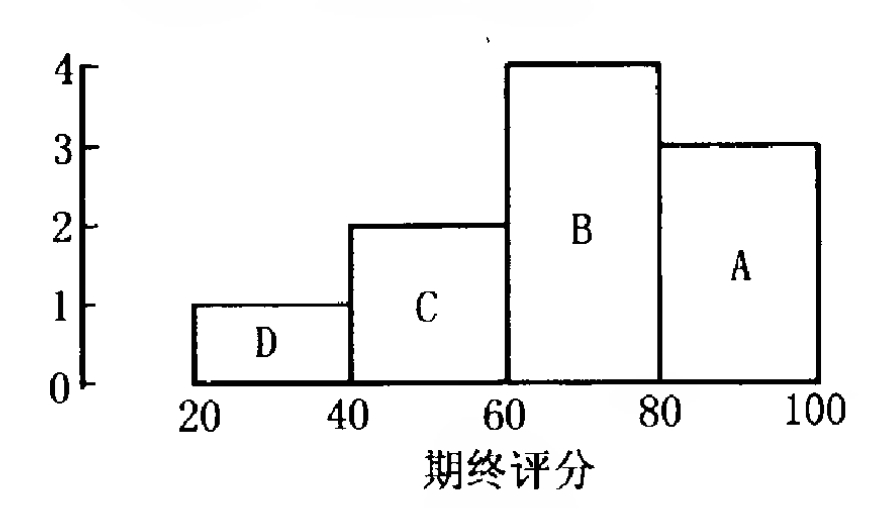
3. 下面的直方图给出了某班期终成绩的分布

(a)有人成绩低于 20 分吗？

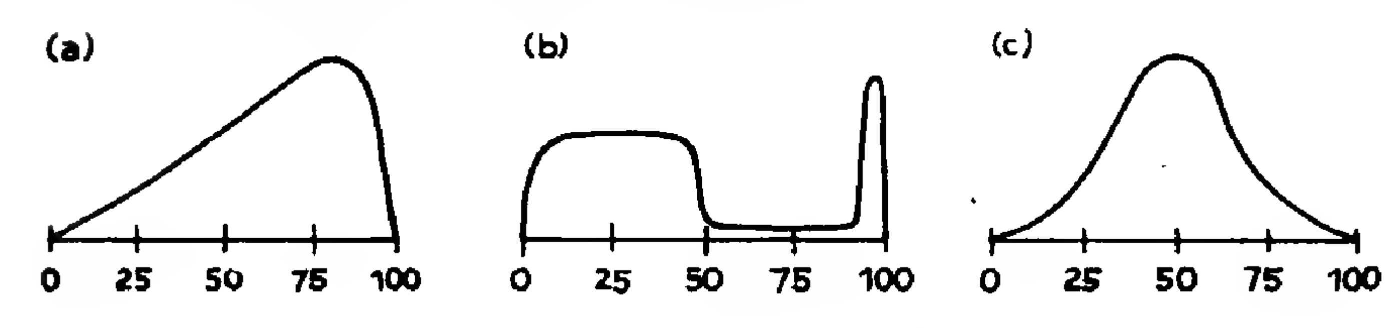
(b)哪个块形表示得分在 60 与 80 分之间的人？

(c)百分之十的人得分在 20 与 40 分之间。大约多少百分数的人得分在 40 与 60 分之间？

(d)大约多少百分数的人得分在 60 分以上？



4. 下面是 3 个不同班级的考试得分直方图的草图。得分是从 0 到 100 分；及格分数是 50。对于每一个班级，及格的百分数大约是 50%呢，还是大大地超过 50%，还是大大地低于 50%？

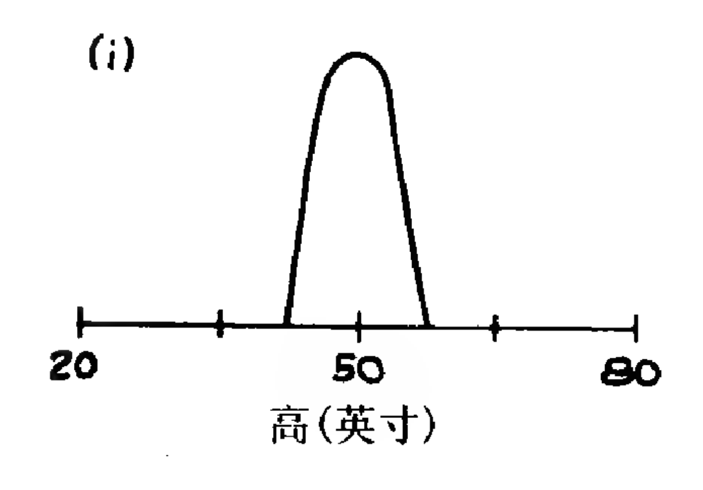


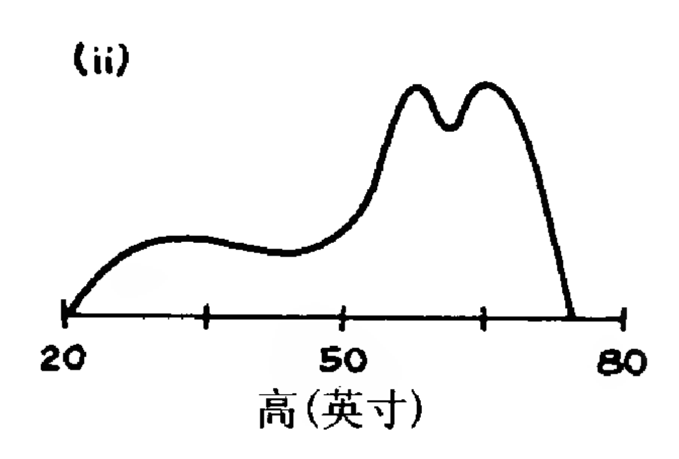
5. 在习题 4 中有一个班级中有两个非常不同的学生组，一组在考试中干得相

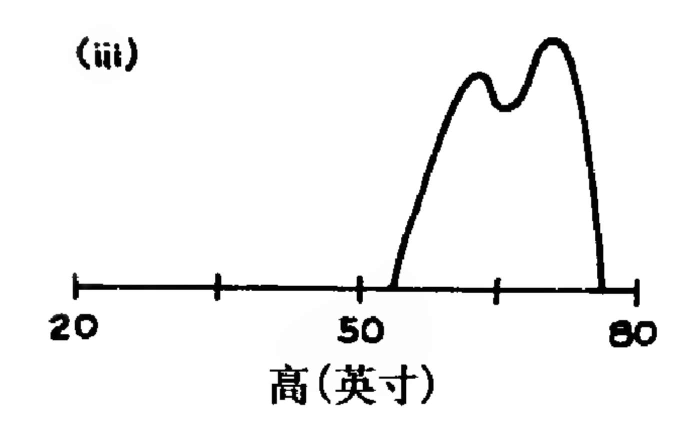
当差，而另一组则非常出色。这是哪一个班级？

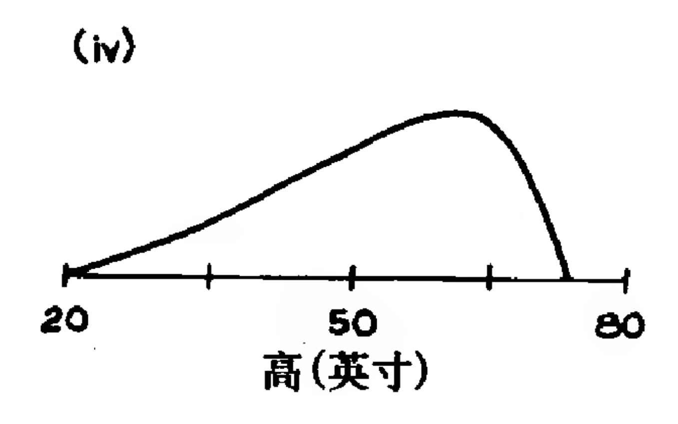
6. 在习题 4 中的(b)班，有较多的人得分在 40—50 范围内呢，还是在 90—100 范围内？

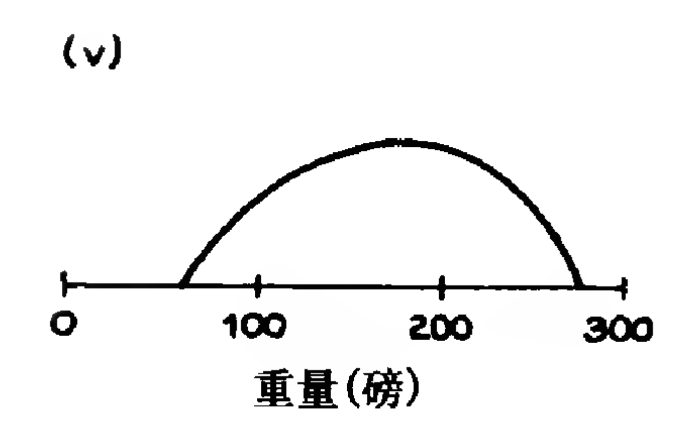
7. 下面给出 6 个草图。其中 4 个是下述(某小镇的研究中的)变量的直方图：

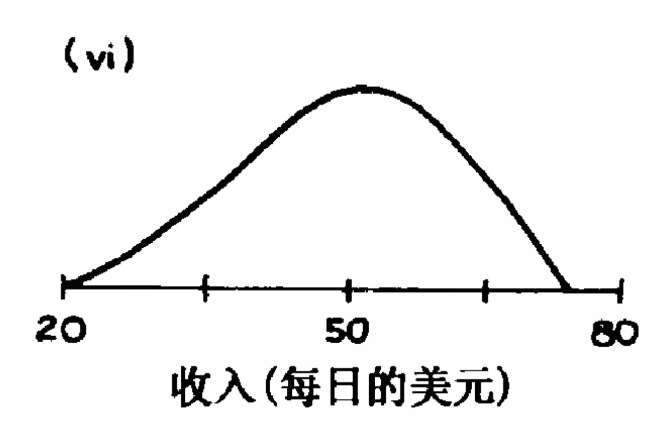












(a)父母双方均小于 24 岁的家庭中所有成员的身高

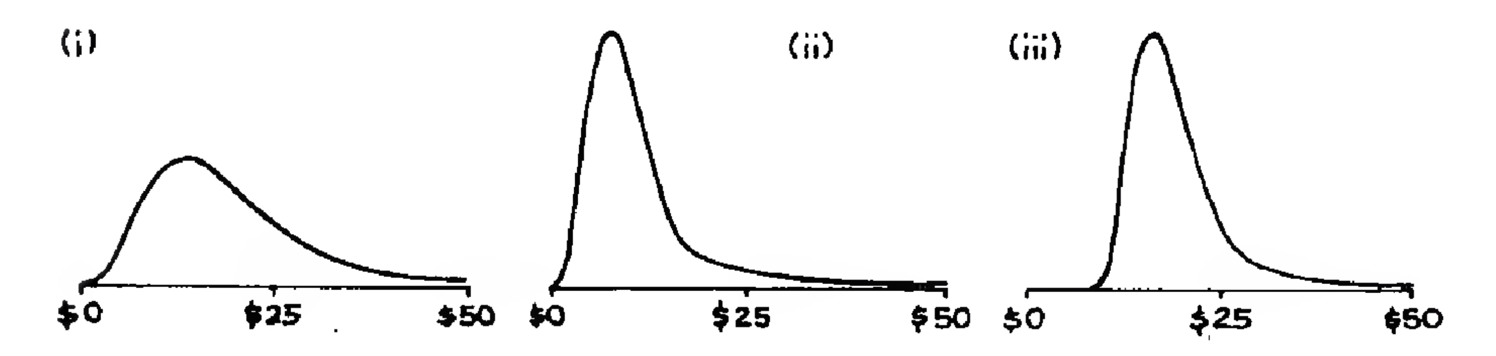
(b)已婚夫妇的身高

(c)全体居民的身高

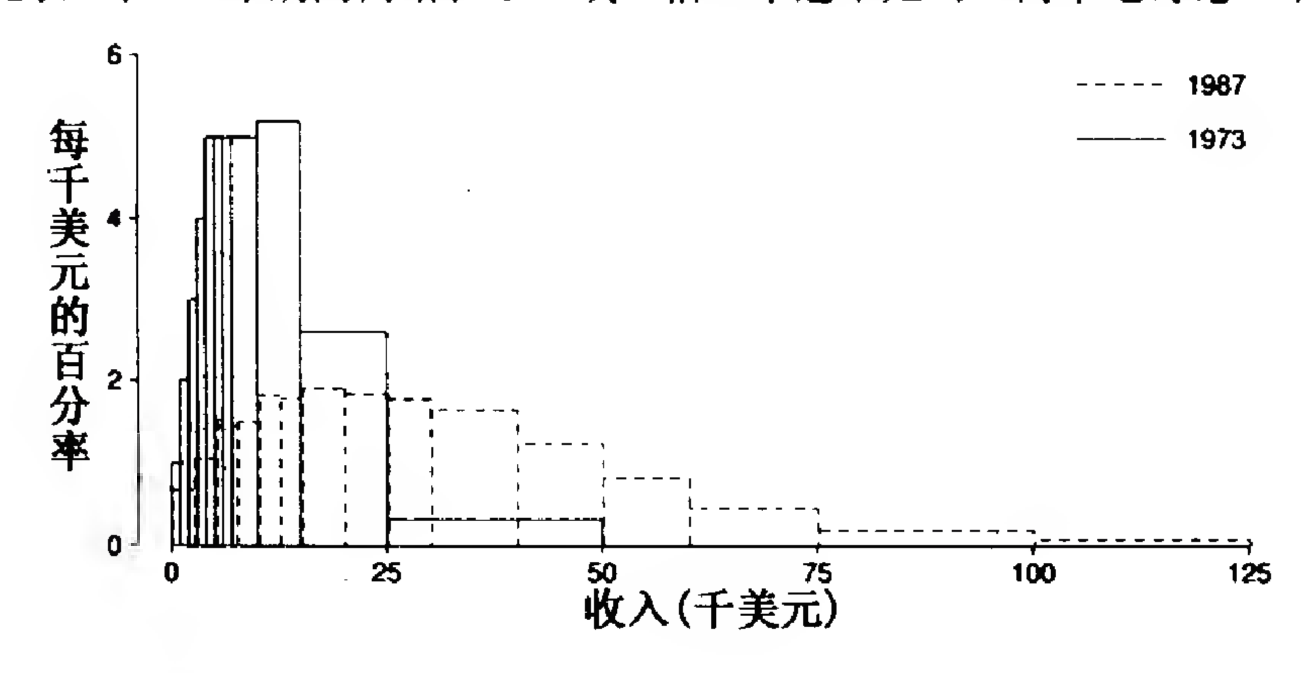
(d)所有汽车的高度

使这些变量和它们的直方图相配。解释你的理由。

8. 一位调研人员收集了 3 组工人按每小时计算的工资的数据。B 组工人所挣得的是 A 组工人的大约 2 倍；C 组工人比 A 组工人每小时多挣了大约 10 美元。哪一个直方图属于哪一组？



9. 下面的图比较了 1973 年与 1987 年美国的家庭收入的直方图。看来仿佛家庭收入在 15 年期间内增长了 2 或 3 倍。难道不是吗？简单地讨论一下。



来源：现场人口调查[3]

这些习题的答案在第 672—673 页上。

2. 绘直方图

本节讲述如何画直方图。方法并不难但有三两个错误要避免。画直方图的起点是分布表，它给出收入在各小组区间内的家庭的百分数(表 1)。这些百分数是通过追溯到原始数据——根据 50 000个家庭——计算求得的。现今这类工作由计算机来做，事实上表 1 就是人口普查局的计算机整理出来的。

对于那些恰好落在两个小组区间之间边界上的家庭，必须告

诉计算机如何处理。这称之为终点约定。表1中遵循的约定由标题指示:左端点包含在小组区间内,右端点不包含在内。例如,在表的第1行中,包含0美元,不包含1 000美元:在这个区间内1个家庭挣0美元或更多,但不到1 000美元。正好挣1 000美元的家庭放进下一个区间。

表1 1973年美国家庭按收入的分布。小组区间包含左端点,但不包含右端点。

收入水平	百分数
$0—$1 000	1
$1 000—$2 000	2
$2 000—$3 000	3
$3 000—$4 000	4
$4 000—$5 000	5
$5 000—$6 000	5
$6 000—$7 000	5
$7 000—$10 000	15
$10 000—$15 000	26
$15 000—$25 000	26
$25 000—$50 000	8
$50 000及以上	1

注:由于四舍五入,百分数加起来不等于100%。

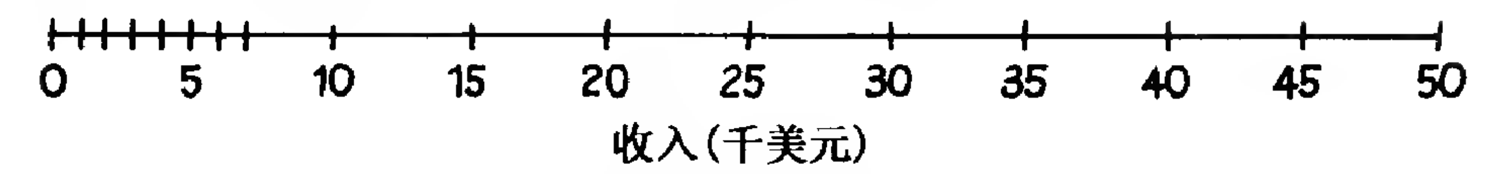
来源:现场人口调查[④]

画直方图的第一步是画一条水平轴。有些人在第一次尝试中得到



这是一个错误。从7 000至10 000美元的区间是6 000至7 000美元区间的3倍长,必须在轴上表示出这一点。另外,对轴进行标

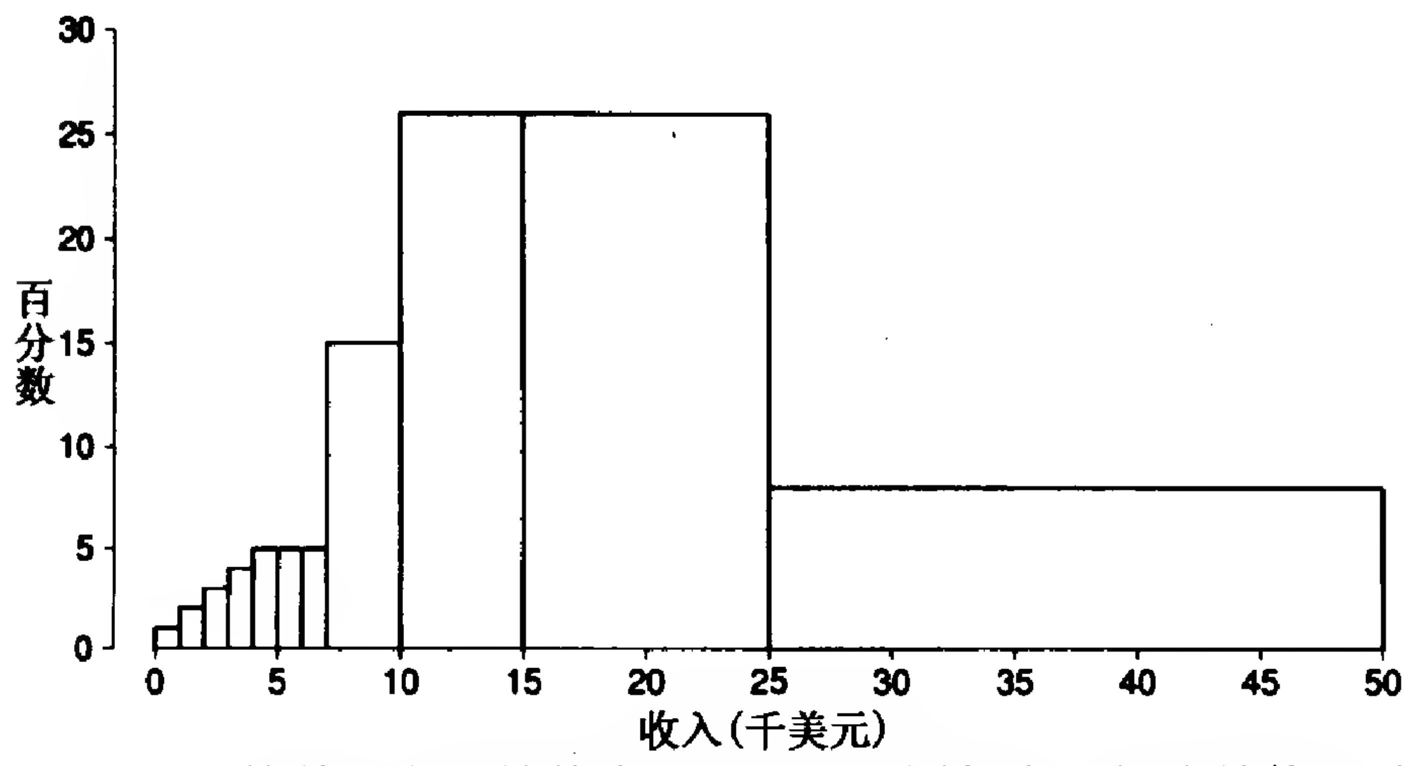
签是一个好主意。因此水平轴看上去应当是这样的：



接下来的一步是画块形。让块形的高等于表中的百分比是很诱人的。图 3 指出如果你犯这个错误的话会发生什么。图像给出收入分布的过份乐观情景。例如，图 3 表明收入超过 25 000 美元的家庭。比收入低于 7 000 美元的多得多。1973 年美国是一个富国，但没那么富裕。

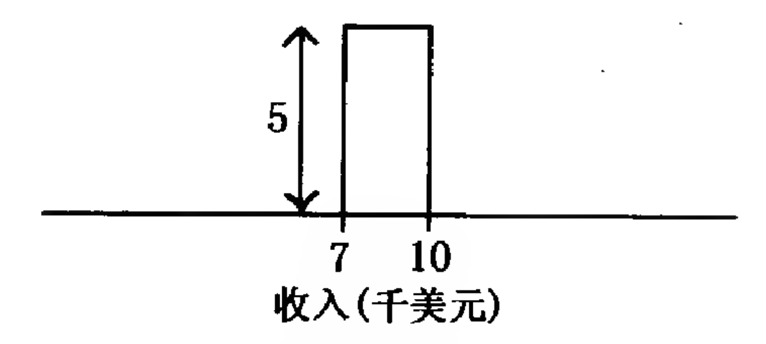
产生困惑的根源在于某些小组区间比其它的要长一些。这意味着表 1 中的百分比不是相互同等的。例如，8%的家庭挣 25 000 至 50 000 美元，比起 15%的挣 7 000 至 10 000 美元的家庭，散布在一个更大的收入范围。直接地标绘百分数忽略了这一点，从而使得在较长的小组区间上的块形过份地大。

图 3　不要标绘百分数。

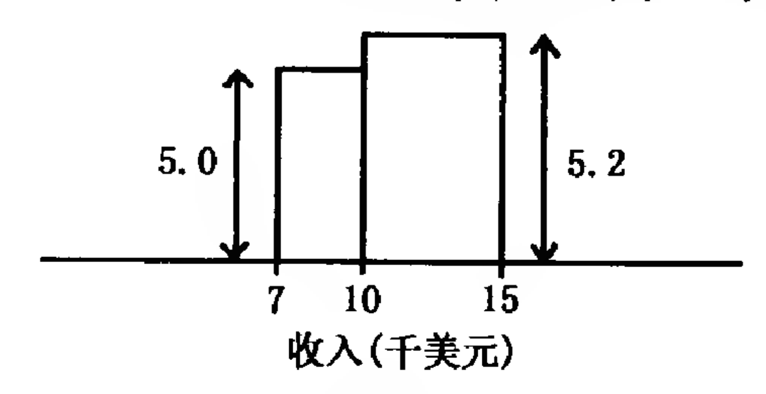


有一个简单的方法补偿小组区间的不同长度。想法是使用千美元区间作为共同的单位。例如，7 000 至 10 000 美元的小组区间包含 3 个这样的区间：7 000 至 8 000 美元，8 000 至 9 000 美元，和 9 000 至 10 000 美元。由表 1，15%的家庭具有这整个区间上的收入。因此，在每一个千美元的子区间内，将只有大约 5%的家庭。

这个 5，而不是 15，就是应当标绘在 7 000 至 10 000 美元这个区间上方的高度。



举第二个例子，取 10 000 至 15 000 美元的区间。它包含了 5 个千美元区间。根据表 1，26%的家庭具有这整个区间上的收入。因此在这 5 个小区间中的每一个内将有大约 5.2%的家庭：26/5=5.2。这给出了 10 000 至 15 000 美元区间上的块形的高度。



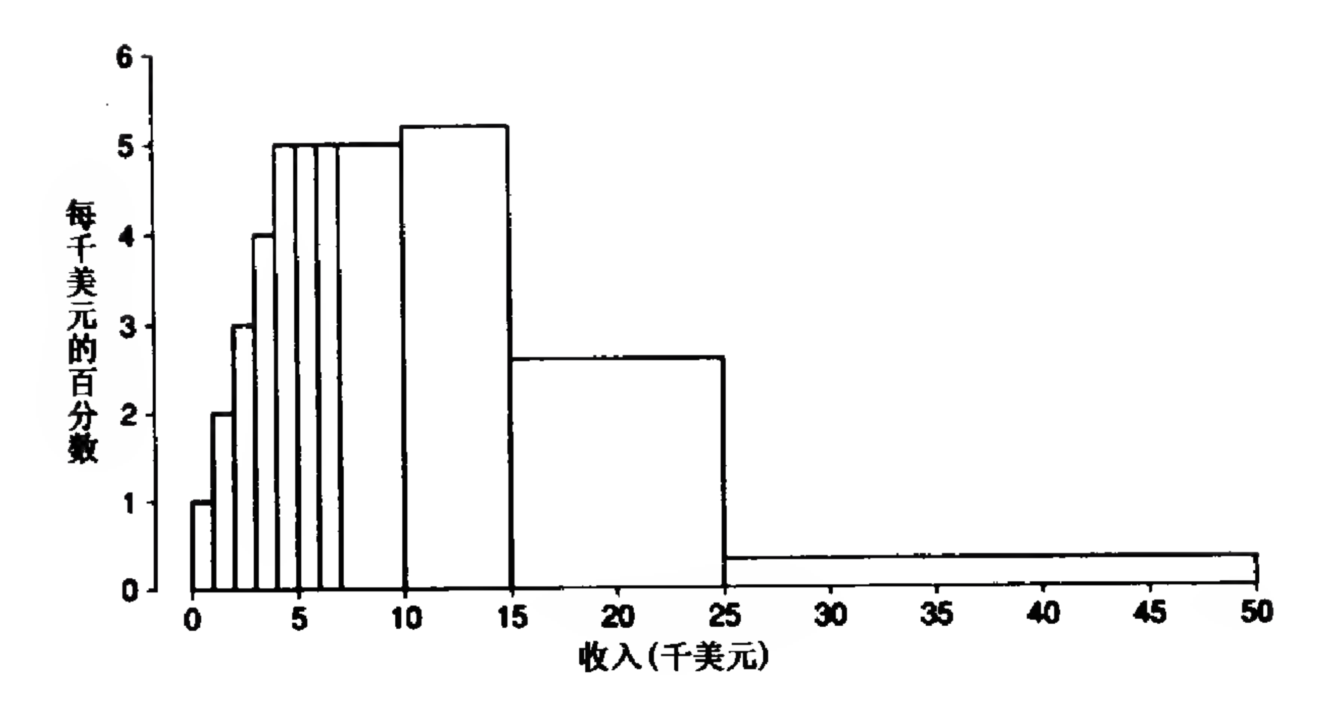
这工作是为表 1 中的两行做的。为了完成直方图，对其余的小组区间做同样的工作。图 4 是结果。

为了计算小组区间上块形的高度，将百分数除以区间的长度。

步骤是简单易懂的，但是纵向刻度的单位有点儿复杂。例如，为了得到 7 000 至 10 000 美元区间上块形高度，用 15 个百分数除以 3 个千美元。这样，答案的单位是每千美元的百分数。考虑“每”的含意恰如你在阅读到东京的每平方英里有 50 000 个人时所理解的：在该城市的每一个平方英里内，有大约 50 000 个人。

关于直方图也是相同的。在 7 000 至 10 000 美元区间上块形的高度是每千美元 5%：在 7 000 至 10 000 美元之间的每一个千美元区间，有大约 5%的家庭。图 4 展示了在纵向刻度上带有这样

图 4　1973 年美国家庭按收入的分布。



的单位的完整直方图。

直方图表示分布,好象百分数在每个小组区间上是均匀地散布的。常常,这是一个好的首次近似。

习题 B

1. 下面的表给出了 1960 年,1970 年,和 1986 年美国 25 岁及 25 岁以上的人的教育水平的分布。"教育水平"是指完成的正规学校教育年数。小组区间包含左端点但不包括右端点;例如,由表的第二行,1960 年大约 14%的人完成了 5—8 年正规学校教育,8 不包含在内;1986 年,大约 5%的人属于这个类。画 1986 年数据的直方图。你可以解释"16 或更多"为 16—17 年正规学校教育;几乎没有人完成 17 年以上的教育。为什么你的直方图在 8,12,和 16 年正规学校教育处有高峰?

教育水平

(正规学校教育年数)	1960	1970	1986
0—5	8	6	3
5—8	14	10	5
8—9	18	13	6
9—12	19	19	12
12—13	25	31	38
13—16	9	11	17
16 或更多	8	11	19

注:由于四舍五人,百分数加起来不等于 100%。

来源:统计摘要,1988 年,表 202

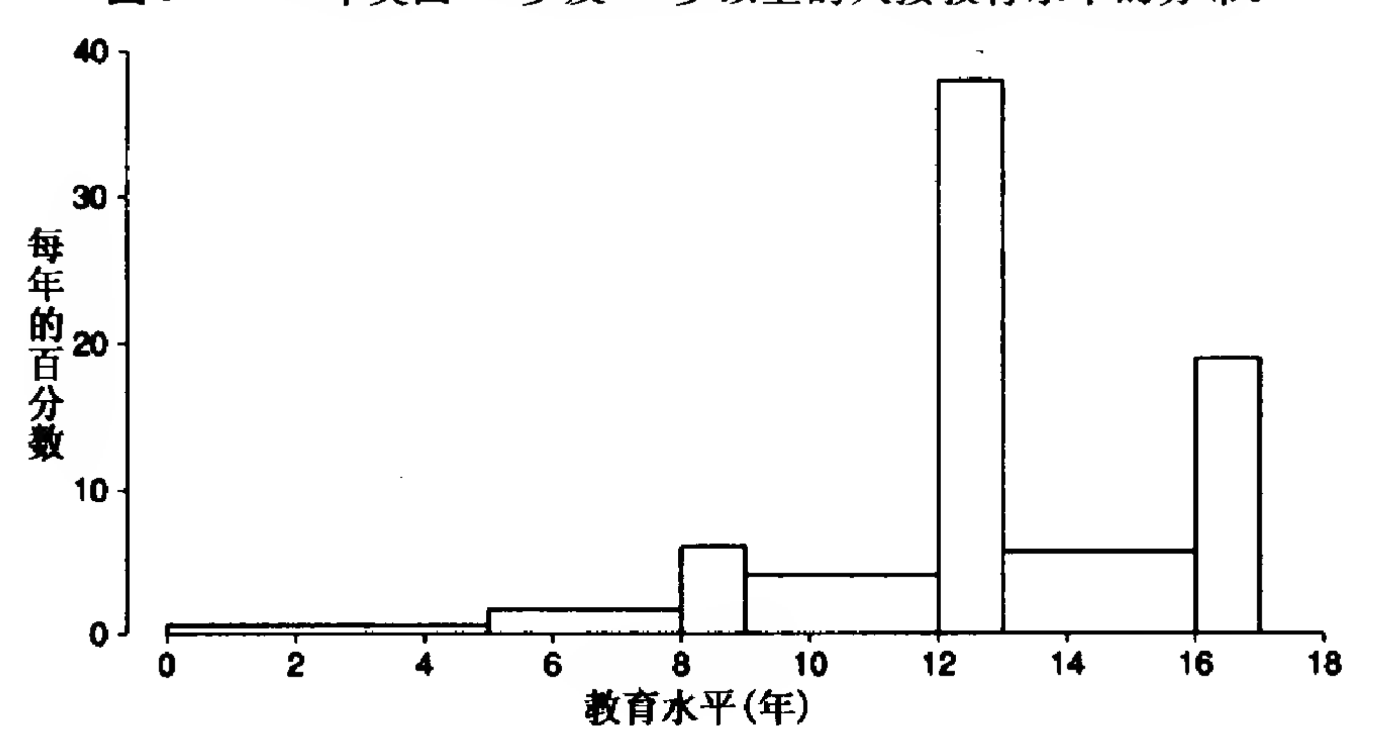
2. 对 1986 年的数据，将头两个小组区间合并成一个(0—8 年，含 8%的人)再画直方图。这样会改变直方图很多吗?
3. 对 1970 年的数据画直方图，并将它与 1986 年的直方图进行比较。在 1970 年与 1986 年之间人口的教育水平发生了什么变化? 是上升，下降，还是停留在大致相同的水平?
4. 从 1960 年到 1970 年教育水平发生了什么变化?

这些习题的答案在第 673 页上。

3. 密度尺度

从直方图上读出面积的时候，有一个纵向尺度会带来方便。在前一节中的收入直方图是利用密度尺度绘制的。水平轴上的单位是 1 000 美元家庭收入。纵轴表示了每 1 000 美元收入的家庭百分数。图 5 是带有密度尺度的直方图的另一个例子。这是 1986 年美国 25 岁以及 25 岁以上的人的教育水平的直方图。"教育水平"是指已完成的正规学校教育年数；幼儿园不算在内。

图 5 1986 年美国 25 岁及 25 岁以上的人按教育水平的分布。



来源：统计摘要，1988，表 202

在这个直方图中所遵循的端点约定有点儿模糊。例如，在 8—9 年区间上的块形表示了已读完了 8 年级但未读完 9 年级的所有人；在 9 年级中途退学的也包含在内。直方图的水平轴上的单位是

年，因此在纵轴上的单位是每年的百分数。譬如在 9－12 年区间上方直方图的高度是每年 4％。换句话说，人口中大约 4％读完了高中第一年，另外 4％读完了高中第 2 年，而另一个 4％则读完了高中第 3 年。

第 1 节描述了直方图中的面积如何表示百分数：如果一个块形比另一个块形包含的面积大些，它表示了那些事例的较大百分数。但是块形的高度表示什么呢？为了回答这个问题，观察图 5 中的水平轴。设想让人排列在该轴上，将每一个人安置在他或她的教育水平上，轴的某些部分——年——将比其他地方拥挤。直方图的高度表示了拥挤程度。

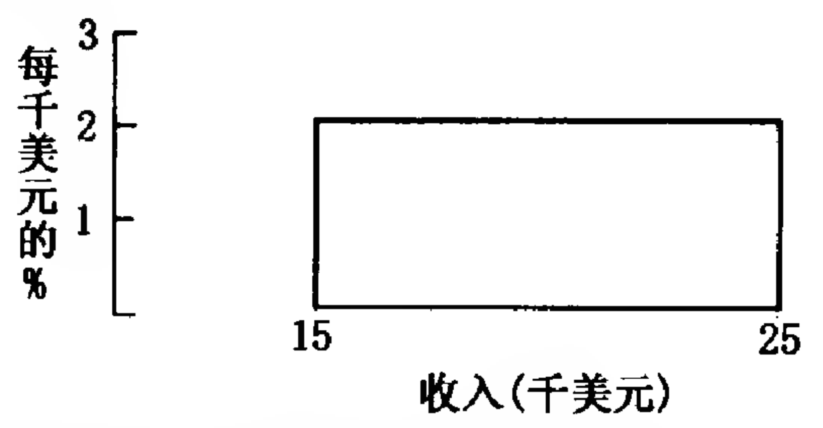
直方图在 12－13 年区间上方为最高，因此在那里拥挤最甚。这个区间包含了所有具有高中程度的人。（在这个区间内的有些人也许继续进过大学，但是他们甚至连大学一年级都没读完。）有两个其它的高峰，一个在 8－9 年（读完小学）和另一个在 16－17 年——读完大学。高峰表明了人们趋向于在三种可能的毕业之一结束自己的学业而不是在这其间中途退学。

最初，也许难于区别用块形的高度所表示的区间上的拥挤这一见解和用块形的面积所表示的区间上的数目。一个例子将讲清楚这点。观察图 5 中 8－9 年区间与 9—12 年区间上面的块形。第一个块形较高，因此这个区间比较拥挤。然而，在 9—12 年区间上的块形有较大一些的面积，因此这个区间包含较多的人。当然，在第二个区间有较多的空间——它是第一个区间的 3 倍长。这两个区间就好象是荷兰与美国。荷兰比较拥挤，而美国拥有较多人口。

一旦你懂得了如何使用密度尺度，在理解直方图中块形的面积时就相当有帮助。例如，取图 5 中 13 至 16 年的区间——这些人读完了大学一年级但是没有毕业。在纵向尺度上，该区间上的块形的高度接近于每年 6％。换句话说，这 3 个一年区间 13—14，14—15，及 15—16 年中的每一个拥有几乎 6％的人。因此整个 3 年区间一定拥有将近 3×6％＝18％的人。1986 年 25 岁及 25 岁以上的

人中几乎 18%读完了大学第一年，但是没有得到学位。

例 1. 下面的草图表示了某城市家庭一收入直方图的一个块形。在该城市大约多少百分数的家庭收入在 15 000 美元至 25 000 美元之间？

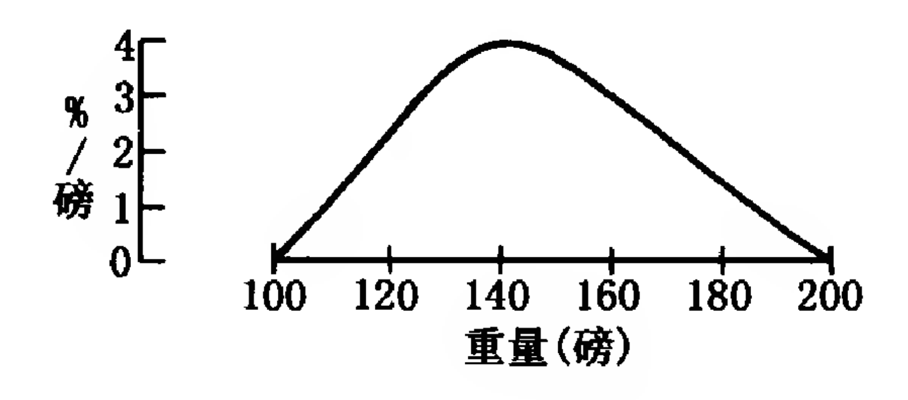


解：区组的高度是每千美元 2%，在 15 000 至 25 000 美元之间的每一个千美元区间包含了该城市大约 2%的家庭。在 15 000 至 25 000 美元之间有 10 个这样的千美元区间。答案是 10×2%＝20%。该城市中大约 20%的家庭收入在 15 000 美元至 25 000 美元之间。

此例说明有了密度尺度，块形的面积以百分数显现。水平单位——千美元——约略了：

2%/每千美元×10 千美元＝20%

例 2 某人使用密度尺度草绘了一些人的体重的直方图。它有什么错误？



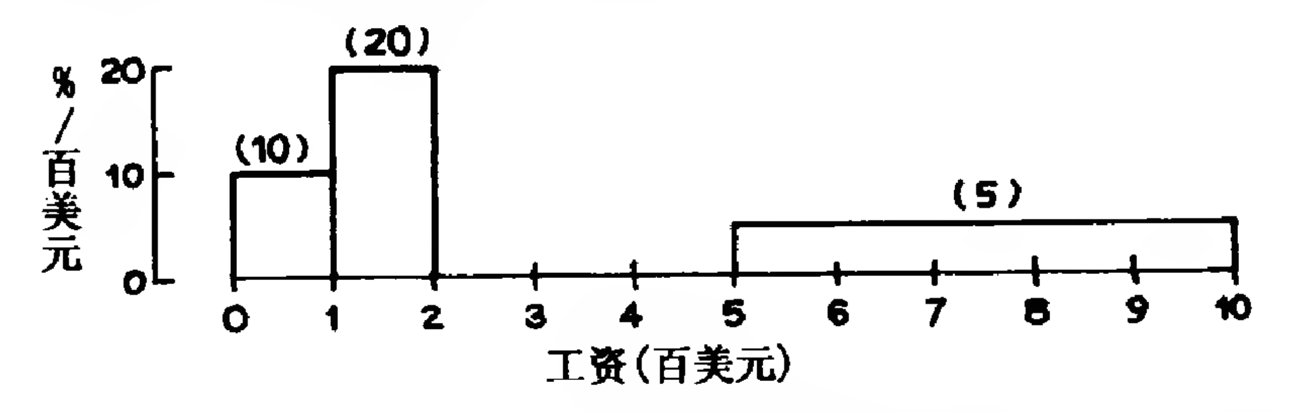
解：总面积是 200%，应当仅为 100%，面积可以计算如下。直方图几乎是一个三角形，它的高是每磅 4%，而它的底是 200 磅－100 磅＝100 磅。面积为

$$\frac{1}{2}\times 底\times 高=\frac{1}{2}\times 100\ 磅\times 4\%/磅=200\%$$

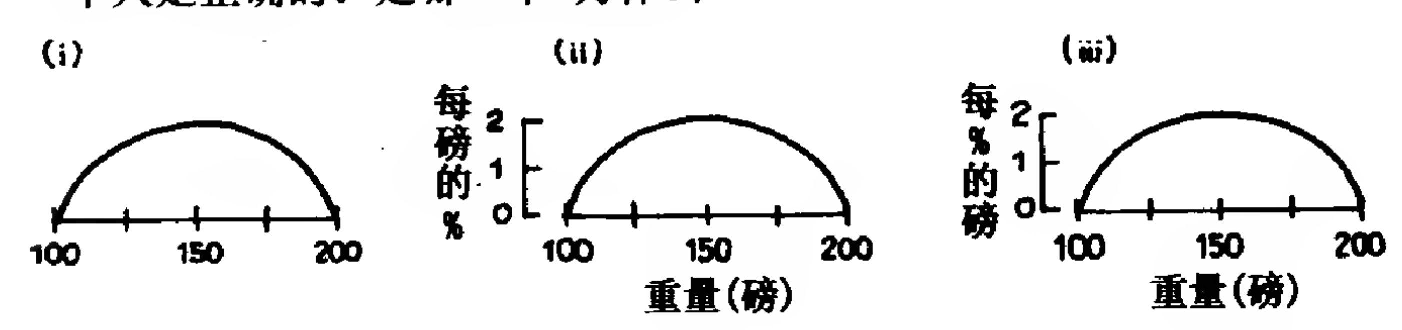
> 在纵轴上带有密度尺度，块形的面积呈现为百分数，因为在水平轴上的单位约略了。一个区间上面直方图下的面积等于该区间内事例的百分数[5]。总面积是100%。

习题 C

1. 下面给出非全日雇员月工资的直方图。没有人一个月挣 1 000 美元以上。200 至 500 美元的小组区间上的块形不见了，它必定有多高？



2. 三个人使用密度尺度对一项研究中的实验对象的体重绘制了直方图。只有一个人是正确的。是哪一个，为什么？



3. 在一项公共卫生总署研究中，绘制了一个直方图显示了每个实验对象(男性现吸烟者)每天所吸烟的支数，如下所示[6]。密度标在圆括号内，小组区间包含右端点而不包含左端点。

   (a)每天吸 10 支烟或不到 10 支烟的人的百分数大约是

   1.5%　15%　30%　50%

   (b)一天吸烟超过一包，但不超过两包的人的百分数大约是

   1.5%　15%　30%　50%

(一包有 20 支烟)

(c)一天吸烟超过一包的人的百分数大约是

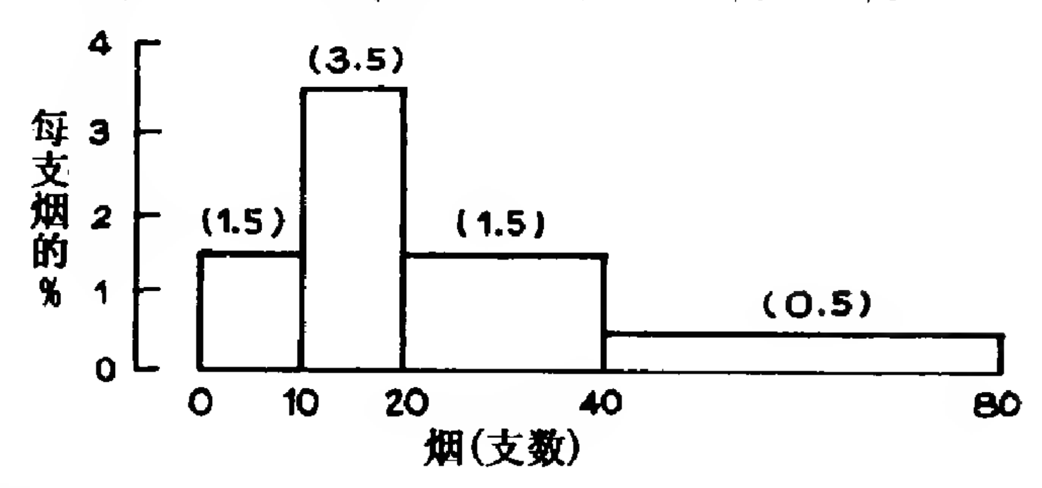
1.5%　15%　30%　50%

(d)一天吸烟超过三包的人的百分数大约是

1%的 0.25　1%的 0.5　10%

(e)每天吸 15 支烟的人的百分数大约是

1%的 0.35　1%的 0.5　1.5%　3.5%　10%



这些习题的答案在第 673 页上。

4. 变量

现场人口调查包括了除收入以外的许多其他变量。变量是在研究中因人不同而变化的一种特性。调查的采访者询问了一组问题:你多大年龄?家里有几口人?家庭的总收入有多少?你结婚了吗?你有工作吗?对应的变量是:年 龄、家庭大小、家庭收入、婚姻状况、和就业状况。有些问题通过给出一个数而得到答案:对应的变量是定量的。年龄、家庭大小和家庭收入是定量变量的范例。有些问题是用修饰语来回答,对应的变量是定性的:例子是婚姻状况(单身、已婚、丧偶、离婚、分居)和就业状况(受雇、失业、不在劳动力之列)。

定量变量可能是离散的或连续的。这不是一个严格且可靠的区分,但很管用⑦。对于一个离散的变量,其值仅可相差确定的量。家庭大小是离散的。两个家庭在大小上可以相差 0 或 1 或 2,等等。在这些量之间不可能有其他量可取。另一方面,年龄是一个连

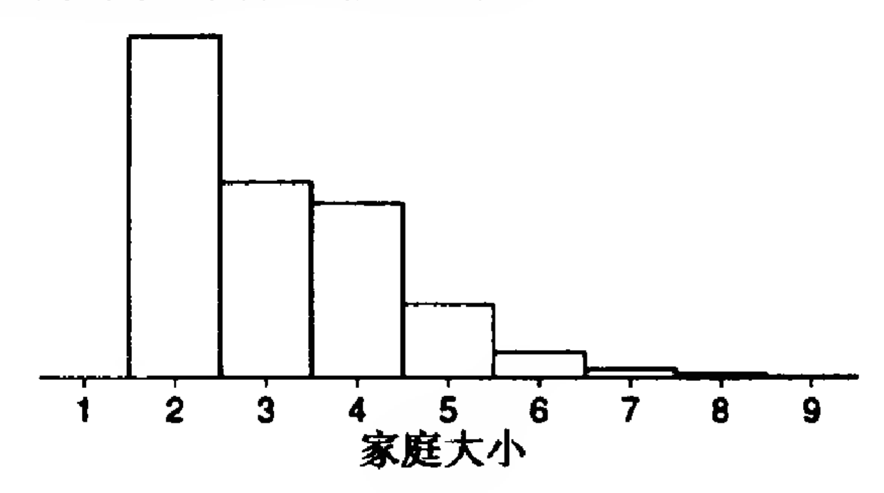
续变量，这并不是由于人是在连续地变老这个事实；它仅仅是指两个人在年龄上的差异可以是任意地小——1 年、1 个月、1 天、1 小时…。最后，定性、定量、离散、和连续这些术语也使用来描述数据——定性数据是在定性变量的基础上收集的，以此类推。

第 2 节讲述了如何从分布表出发绘制直方图。常常出发点是原始数据——一列事例（个体、家庭、学校等）和对应的变量值。为了画直方图，必须准备一张分布表。第一步是选择小组区间。具有过多或过少的小组区间，直方图将不会提供信息的。没有什么规则，这是一件判断或尝试和误差的事情。通常是从 10 或 15 个小组区间出发并由此展开工作。在本书中，小组区间将总是给定的[⑧]。

在对一个连续变量绘制直方图时，你也必须确定终点约定；就是说，对于正好落在边界上的事例如何处置。关于离散变量，有一个约定避开了这种麻烦：小组区间的中心定在变量的值上。例如，家庭大小可以是 2 或 3 或 4，等等（人口调查不认为一个人是一个家庭。）分布表中对应的小组区间为

中心	小组区间
2	1.5 至 2.5
3	2.5 至 3.5
4	3.5 至 4.5
⋮	⋮

图 6 直方图表示了 1988 年家庭大小的分布。对于一个离散的变量，小组区间的中心定在可能取的值上。



来源：现场人口调查数据磁带。

由于一个家庭不可能有 2.5 个成员，有关 2.5 属于哪个区间就不

成问题。图 6 绘出了家庭大小的直方图。

习题 D

1. 将下面的变量中的每一个归类为定性的或定量的，如果是定量的，归类为离散的或连续的。

   (a)职业 (b)居住区域 (c)体重

   (d)身高 (e)拥有的汽车数

2. 在 1960 年，继而在 1980 年，美国妇女被询问："你有了几个孩子？"结果如下所示。

   (a)变量是离散的还是连续的？

   (b)画这些数据的直方图(你可以取"9 或更多"为 9——几乎没有妇女有超过 9 个的孩子)。

   (c)你的结论是什么？

18 岁及 18 岁以上的妇女按已养育过的小孩数的分布

孩子数	妇女的百分数(1960)	妇女的百分数(1980)
0	22	29
1	17	16
2	21	22
3	16	15
4	10	8
5	5	4
6	3	2
7	2	1
8	2	1
9 或更多	3	1

注：由于四舍五入，百分数加起来不等于 100%。

来源：(i)1960 年，(美国)国家卫生统计中心提供的 HES 数据磁带。(ii)1980 年，人口普查局提供的人口调查数据磁带。

这些习题的答案在第 674 页上。

5. 对变量的控制

六十年代，许多妇女开始使用口服避孕药物，俗称"药片"。由

于口服避孕药改变了人体内分泌平衡，因而重要的是看看有什么副作用。这个问题的研究由设在加利福尼亚(California)州 Walnut Creek 的 Kaiser 诊所里的避孕药研究项目承担。在 Walnut Creek 地区超过 20 000 名妇女属于 Kaiser 基金会健康计划，该计划为她们支付每月的保险费用并从 Kaiser 获取医药服务。这些服务的内容之一是一项称为"多相"的常规检查。在 1969—1971 年期间，大约 17 500 名 17—58 岁的妇女接受多相检查从而成为药物研究项目的对象。

研究人员们比较了两个不同的妇女组的多相结果：

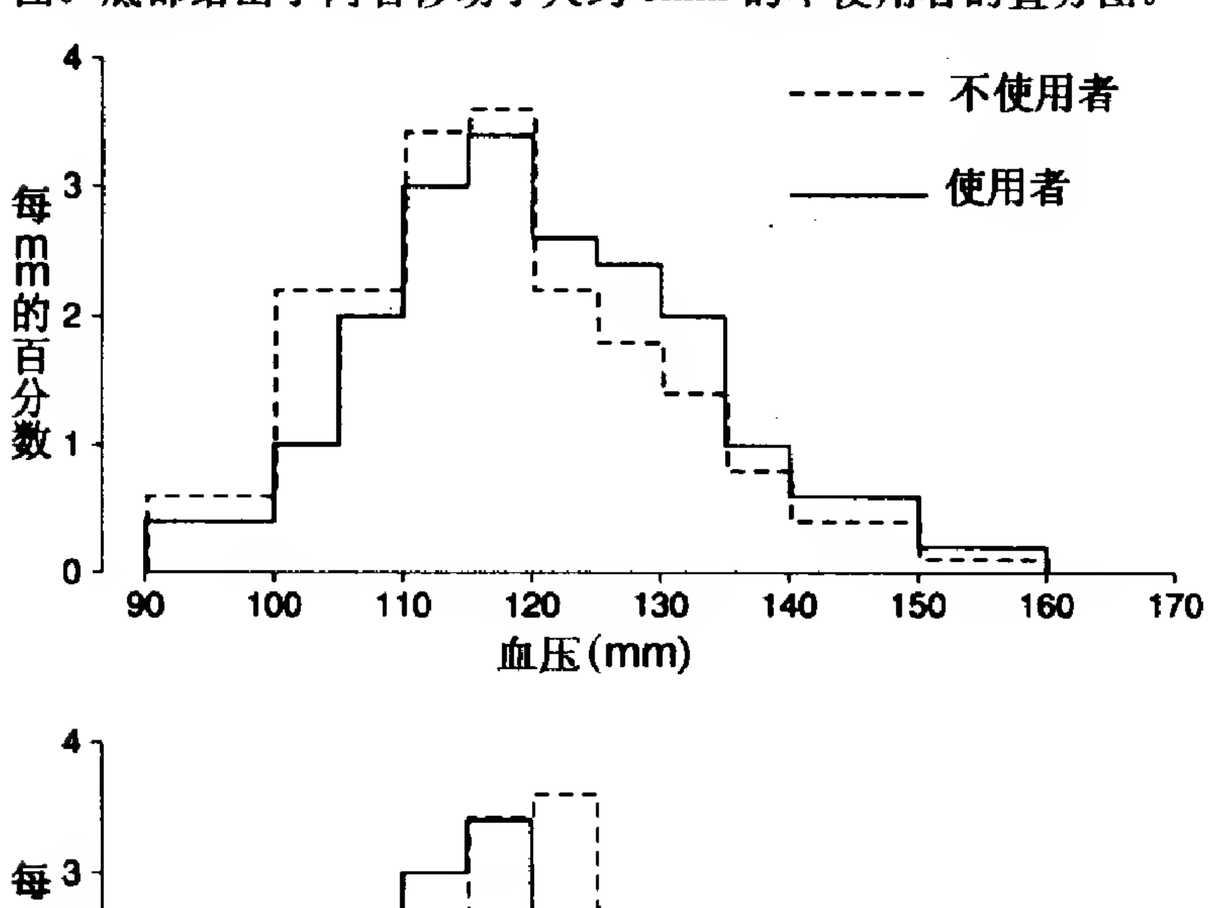
- "使用者"使用口服避孕药(试验组)；
- "不使用者"不使用口服避孕药(对照组)。

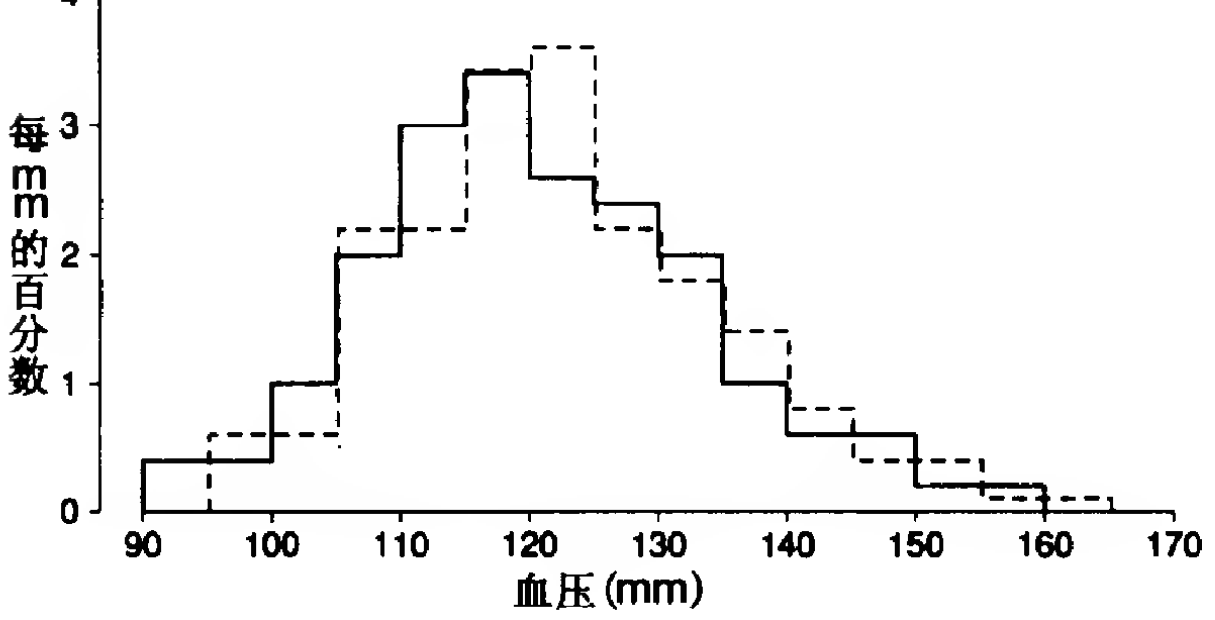
这是一项观察研究。因为是妇女本人决定是否服用口服避孕药，研究人员只是观察所发生的事情。

一个问题是口服避孕药对于血压的影响。也许很自然地去比较使用者与不使用者的血压直方图。然而，这可能会引入岐途。血压随着年龄趋于上升，而整体来说不使用者比使用者较年长一些。例如，与 50%的使用者超过 30 岁相比较，大约有 70%的不使用者超过 30 岁。年龄的效应与口服避孕药的效应混杂在一起。为了使口服避孕药物的全部效应能显露出来，必须对每一个年龄组作各别的比较：这就控制了年龄[9]。比较结果相当地类似，因此我们将只观察 25—34 岁的妇女。图 7 给出了该年龄组中使用者与不使用者的直方图。

图 7 的顶部中的两个直方图有着非常相象的形状。但是，使用者的直方图在 120mm 的右边较高一些而在左边则较低一些。高血压(120mm 以上)在使用者中间更普遍一些，而低血压较不普遍。现在设想将 5mm 加到每一个不使用者的血压上去。那将使他们的直方图如同图 7 底部所示那样向右移动 5mm。在底部的图中，两个直方图匹配得相当好。就所涉及的直方图，仿佛是使用了口服避孕药对每位妇女的血压加了大约 5mm。

图 7 口服避孕药的影响。顶部给出了避孕药研究项目中 25-34 岁妇女中 1 747 位使用者与 3 040 位不使用者的收缩血压的直方图。底部给出了向右移动了大约 5mm 的不使用者的直方图。

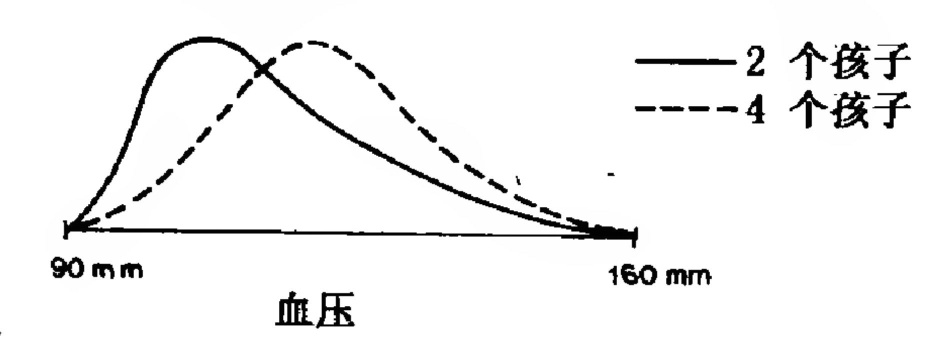




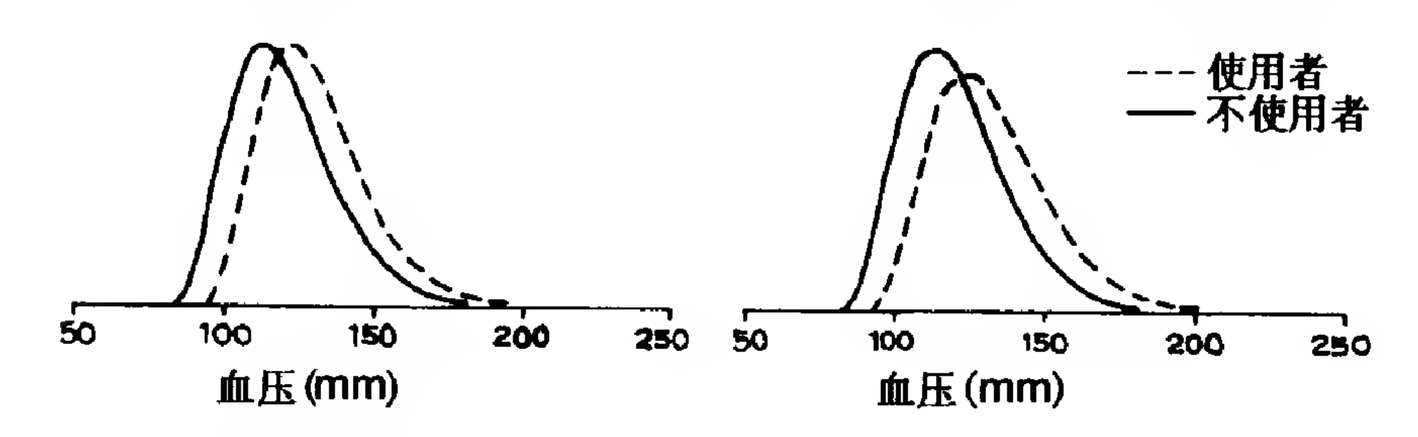
这个结论必须谨慎对待。避孕药研究项目的结果提出了假如一位妇女继续使用口服避孕药，她的血压将上升大约 5mm。但是证据是不完全的。由于设计的原因不可能完全。药物研究项目是一项观察研究而不是对照实验。本书第一部分指出了观察研究关于因果关系可能会出现误导。可能存在一些既不是口服避孕药又不是年龄那样的尚未查明的因素，它们在影响着血压。对于药物研究项目来说，这就有点儿离题了。口服避孕药影响血压的生理原理已经完全建立。药物研究项目揭示了这种影响的大小。

习题 E

1. 作为一种额外的收获，药物研究项目对拥有不同的孩子数的妇女比较了血压，下面是关于拥有 2 个或 4 个孩子的妇女的直方图的草图。哪一组有较高的血压？有孩子会引起母亲的血压改变吗？还是这种变化可能归因于某些其他因素，它们的效应与拥有孩子效应混杂在一起？



2.（假设的）这里草图表示了对于 25—29 岁的妇女关于口服避孕药的两项其他研究的结果。在一项研究中，口服避孕药增加血压大约 10mm；在另一项研究中，口服液孕药增加血压 10%左右。哪一个图配哪一项研究，为什么？



这些习题的答案在第 674 页上。

6. 交叉列表

前面一节讲述了如何控制年龄的影响：这是一种对每一个年龄组各别做比较的事情。这种比较通过图 7 中的直方图形象化地给出。一些调查研究人员喜欢采用一种叫做交叉列表的方法以表的形式来进行比较。按年龄和口服避孕药使用的关于血压的交叉列表在表 2 中给出。这样的表给人有点儿强制的味道，眼睛很自然地扫过它们直到某些所需要的数字为止。整个表 2 相当于对每一个年龄组的使用者和不使用者各别作的血压分布表。

表 2 对于避孕药研究项目中的妇女，不包括那些怀孕的以及服用不同于口服避孕药的内分泌药物的人，按年龄和口服避孕药使用的收缩血压。小组区间包括左端点，不包括右端点；'—'表示是可以忽略的。由于四舍五入的缘故百分数加起来可能达不到100%。

血压 (mm)	17—24 岁 不使用者	17—24 岁 使用者	25—34 不使用者	25—34 使用者	35—44 不使用者	35—44 使用者	45—58 不使用者	45—58 使用者
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
90 以下	—	1	1	—	1	1	1	—
90—95	1	—	1	—	2	1	1	1
95—100	3	1	5	4	5	4	4	2
100—105	10	6	11	5	9	5	6	4
105—110	11	9	11	10	11	7	7	7
110—115	15	12	17	15	15	12	11	10
115—120	20	16	18	17	16	14	12	9
120—125	13	14	11	13	9	11	9	8
125—130	10	14	9	12	10	11	11	11
130—135	8	12	7	10	8	10	10	9
135—140	4	6	4	5	5	7	8	8
140—145	3	4	2	4	4	6	7	9
145—150	2	2	2	2	2	5	7	9
150—155	—	1	1	1	1	3	2	4
155—160	—	—	—	1	1	1	1	3
160 及以下	—	—	—	—	1	2	2	5
总百分数	100	98	100	99	100	100	99	99
总　数	1 206	1 024	3 040	1 747	3 494	1 028	2 172	437

考察 17—24 年龄组的一列。有 1 024 位使用者和 1 206 位不使用者。大约 1%的使用者血压低于 90mm；不使用者中相对应的百分数是微不足道的——那就是表中破折号所表示的意思。(每列的百分数应当总起来达到 100，但是由于四舍五入的缘故可能相差 1%左右。)为了知道口服避孕药对于 17—24 岁的妇女的血压

的影响。要紧的事情是观察17—24岁组中使用者与不使用者在列中的百分数。为了知道年龄的影响，首先观察在每个年龄组中不使用者的这一列并且看，随着年龄的上升百分数如何向高血压转移。然后对使用者做同样的事情。

习题 F

1. 利用表2中的交叉列表，回答下列问题：
   (a)多少百分数的17—24岁的使用者具有140mm或更高的血压？
   (b)多少百分数的17—24岁的不使用者具有140mm或更高的血压。
   (c)你得出什么结论？
2. 对17—24岁的使用者与不使用者的血压绘制直方图。你得出什么结论？
3. 比较17—24岁的不使用者与25—34岁的不使用者的血压直方图。你得出什么结论？

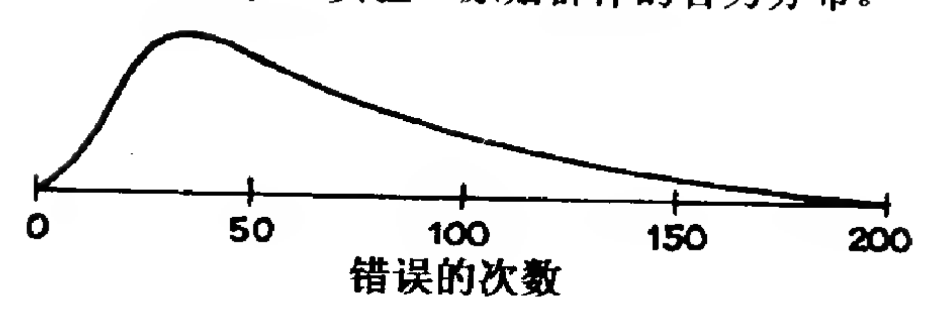
这些习题的答案在第674页上。

7. 选择性繁殖

1927年，心理学家Charles Spearman出版了他的有关人类智力理论的著作：The Abilities of Man（人的能力）。简要地说，Spearman认为智力的能力（诸如阅读理解，算术，或者空间想象）的测验分数是两个独立分量的加权和：一个是被Spearman称为"g"的一般智力因素，和另一个特定于每次测验的能力因素。这个理论引起了众多注意。

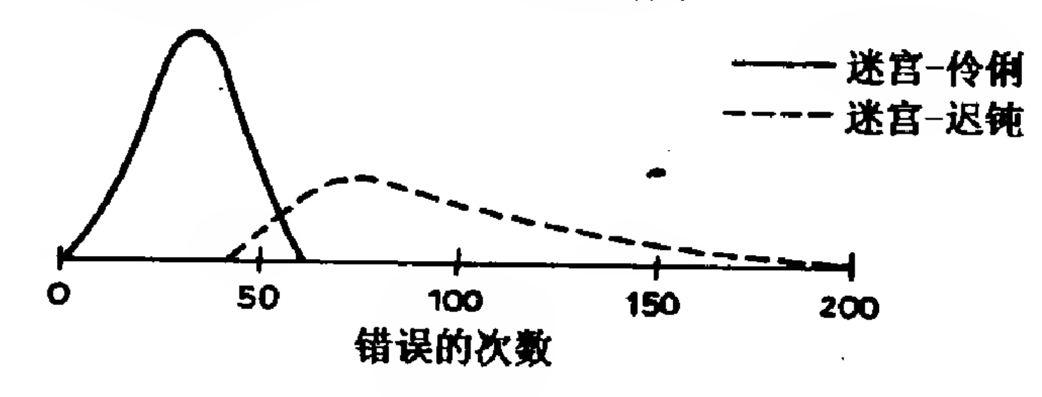
作为Berkeley（伯克莱）大学心理系Ph. D研究的一部分，Robert Tryon决定对一个控制外界变量比较简单的动物群体检验这个理论[10]。Tryon选用了在实验室中容易繁殖的老鼠。为了测验老鼠的智力，Tryon将它们放入一个迷宫。当老鼠在迷宫中奔跑时，由于走进了死胡同而犯了错误。测验由19次走迷宫组成；动物的"智力分"是所犯错误的总次数。因此伶俐的老鼠是那些得低分的，而迟钝的老鼠则是那些得高分的。Tryon从142只老鼠着手进行，它们的智力分的分布草拟在图8。

图 8 Tryon 实验。原始群体的智力分布。



实验的下面一步是按智力繁殖。在每一代中，"迷宫一伶俐"老鼠（仅仅犯少数错误的那些）互相繁殖。同样地，"迷宫一迟钝"动物（具有高分的）在一起繁殖。七代以后，Tryon 有了 85 只迷宫一伶俐品系的老鼠和 68 只迷宫一迟钝品系的老鼠。它们之间在分数上有明显的间隔。图 9 给出了这两个组的智力分布，它们的直方图几乎没有交迭（事实上，Tryon 在七代之后继续进行选择性繁殖，但是在分数上没有得到更多的分离。）。

图 9 Tryon 实验，在选择性繁殖七代以后，在迷宫一伶俐"与"迷宫一迟钝"品系之间存在一个明显的间隔。



由 Tryon 造就的两个品系仍为伯克莱大学心理系所使用（用于其他的实验）。许多代以后，来自迷宫一伶俐总体的老鼠继续在走迷宫中比来自迷宫一迟钝总体的老鼠表现出色。

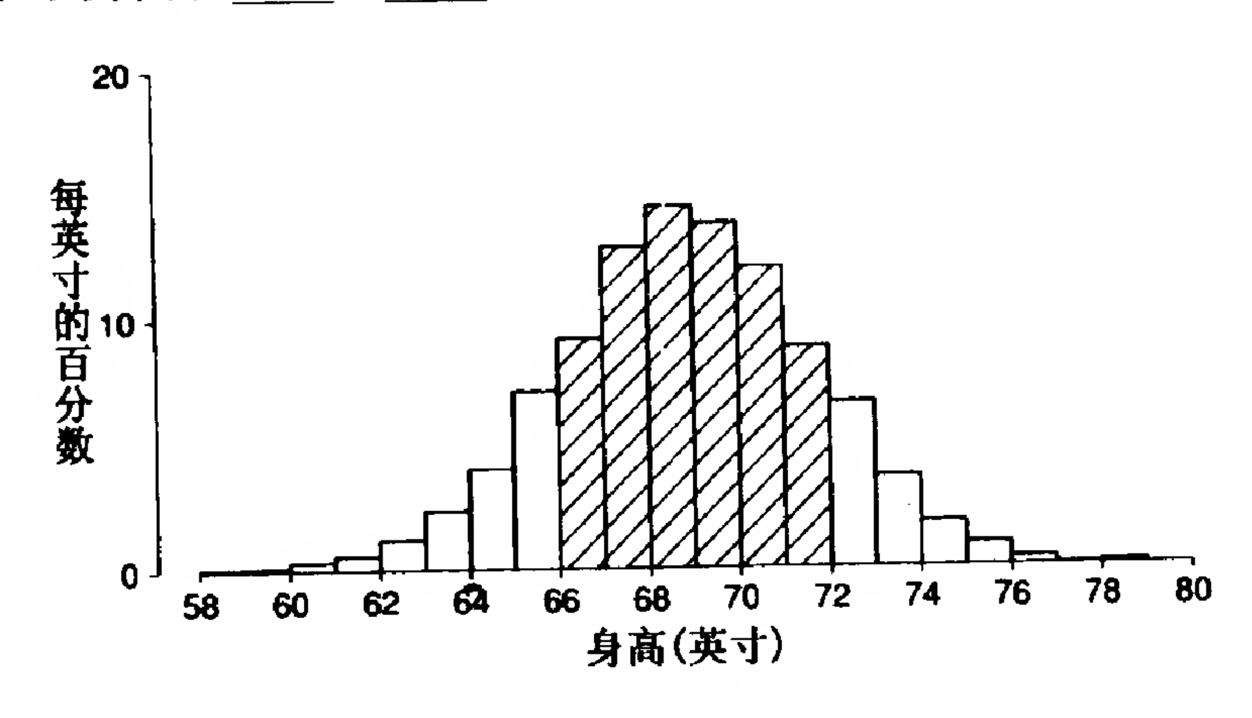
这样 Tryon 为了智力能力的目的而操纵繁殖——显示出某些智力能力至少部分地从一开始就已确定。这个实验对于 Spearman 的理论说明了什么？Tryon 发现在其他的动物智力测验方面，诸如几何图形之间或者光的强弱之间的判别等，"迷宫一伶俐"老鼠做得并不好于迷宫一迟钝的老鼠。这是反对 Spearman 的一般智力因素理论的一个强有力的证据（至少对于老鼠）。

另一方面，Tryon 发现了在这两个老鼠群体之间引人兴趣的心理差异。“伶俐者”看来是不善交际的性格内向者，对于迷宫中的生活能很好地调整，但是在与其它老鼠的关系中有点神经质，而“迟钝者”则完全相反。

8. 复习题

复习题可能包含了前面几章的内容。

1. 下面给出了一组有代表性的男性样本的身高直方图。阴影面积表示了身高在____和____之间的男人的百分数。



来源：关于政治与社会研究的校际合作所提供的数据磁带。

2. 1986 年在美国按年龄的分布如下所示。绘制直方图。(小组区间包含左端点，不包含右端点；例如，由表中的第二行，13％的人年龄为 5 岁及以上但不到 14 岁。“75 及以上”区间可以终止于 85 岁，男女均包含在资料中。)使用你的直方图回答下列问题。

(a)30 岁及以上的人的百分数大约是 25％，50％，还是 75％？

(b)总体中 18 岁的百分数大约是 2％，5％，还是 8％？

年龄	占人口的百分数	年龄	占人口的百分数
0—5	8	35—45	14
5—14	13	45—55	9
14—18	6	55—65	9
18—22	6	65—75	7
22—25	5	75 及以上	5
25—35	18		

来源：统计摘要，1988，表 13。

3. 来自 1980 年人口普查的数据可以用来发现按房间数的居住房单位（包括公寓房间）的分布。人口调查对"私房"和"租房"单元分别进行该项工作。纽约市的结果如下所示。

(a) 对这两个分布中的每一个绘制直方图（你可以假定"8 或更多"意思是指 8 或者 9；极少数单元有 9 间以上的房间——尤其是在纽约。）

(b) 租住的百分数相加达 100.1%。为什么？

(c) 对于租住的房子，1 房间单元的百分数较大一些。那是因为在总数上有较多的租住房吗？回答是或否，并作简短的解释。

单元中的房间数	私房（百分数）	租房（百分数）
1	0.6	7.5
2	1.2	10.6
3	6.7	33.2
4	14.0	28.5
5	20.9	13.5
6	29.9	5.3
7	14.1	1.0
8 或更多	12.6	0.5
总计	100.0	100.1
总数	652 105	2 136 425

来源：住房人口调查，1980

4. 下面的图是一个直方图，表示了在药物研究项目（第 5 节）中所有 14 148 位妇女的血压分布。使用直方图回答下列问题：

(a) 血压在 130mm 以上的妇女的百分数大约是 25%，50%，还

是 75%。

(b) 血压在 90mm 与 160mm 之间的妇女的百分数大约是 1%，50%，还是 99%？

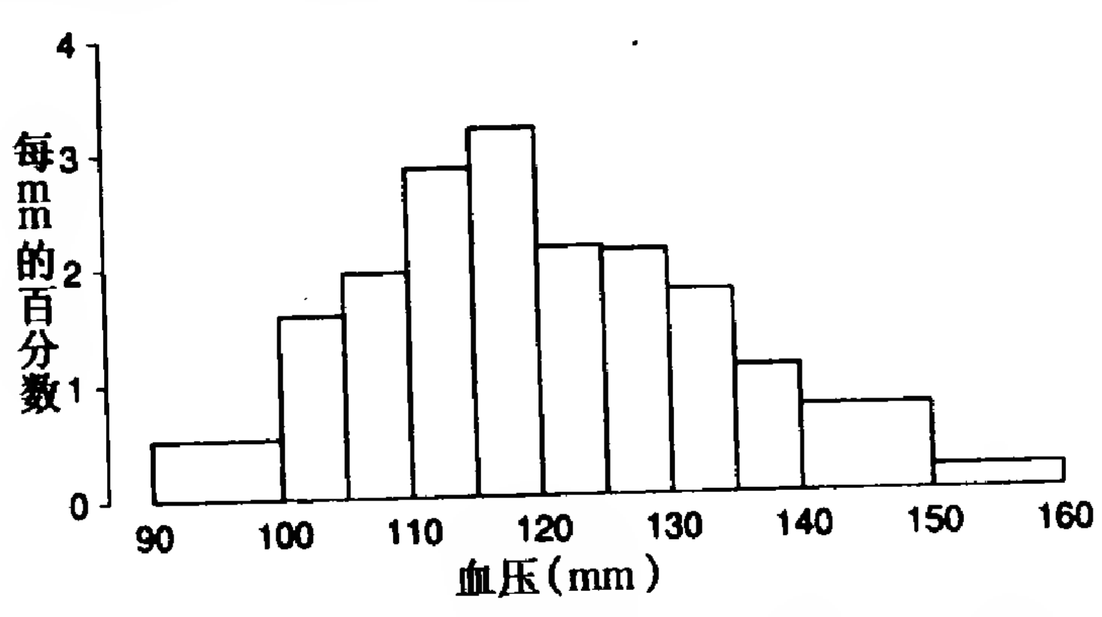
(c) 在哪个区间有较多妇女：135—140 还是 140—150mm？

(d) 哪个区间更拥挤一些：是 135—140mm 还是 140—150mm？

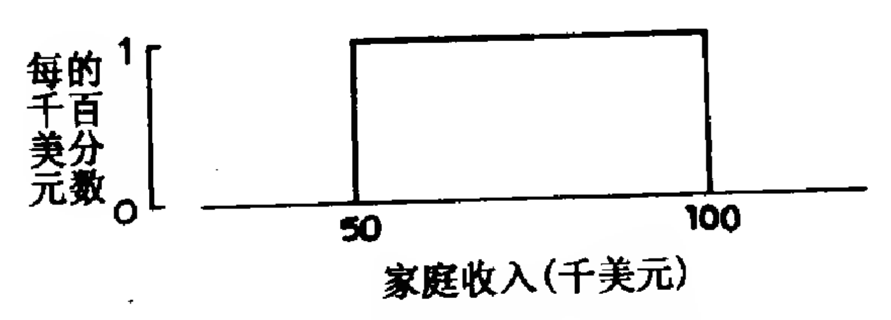
(e) 在区间 125—130mm 内，直方图的高大约为每 mm2.1%。多少百分数的妇女具有这个小组区间里的血压？

(f) 哪个区间有较多妇女：是 97—98mm，还是 102—103mm？

(g) 所有的毫米中，哪里最拥挤？

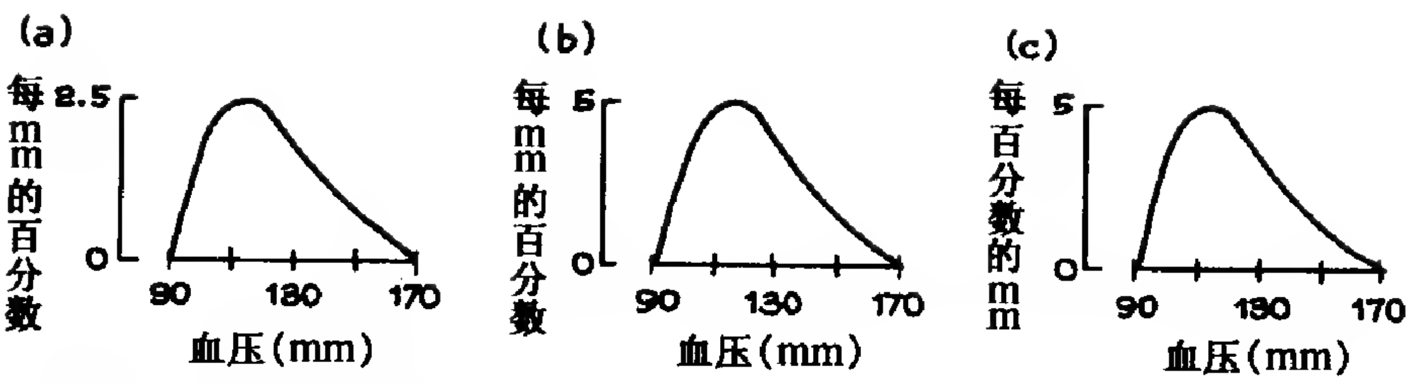


5. 有人草绘了一个富裕郊区家庭收入直方图的一个块形。在该区中大约多少百分数的家庭一年的收入在 90 000 美元和 100 000 美元之间？



6. 三人使用密度尺度试图对某项研究中对象的血压草绘直方图。只有一个人是正确的。是哪一个人，为什么？(见下页图 a、b、c)

7. (假设的)在一项研究中，100 个人的身高被测量到最接近的 1/8 英寸。有关结果的直方图如下所示。下面目录中的两个有该直

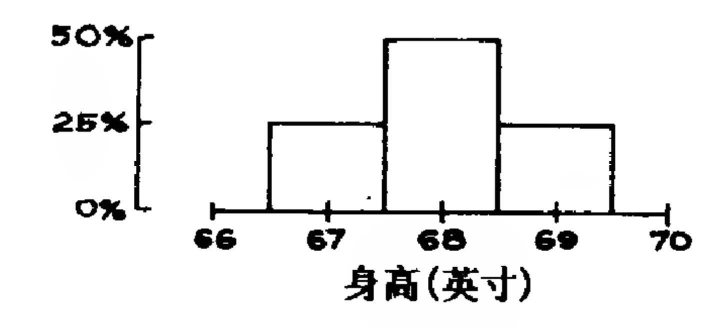


方图。哪两个，为什么？

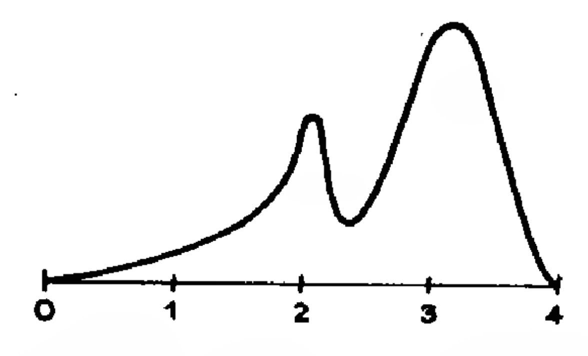
(i)25 人，67 英寸高；50 人，68 英寸高；25 人，69 英寸高。

(ii)10 人，66¾英寸高；15 人，67¼英寸高；50 人，68 英寸高；25 人，69 英寸高。

(iii)30 人，67 英寸高；40 人，68 英寸高；30 人，69 英寸高。



8. 在一项由加利福尼亚—伯克莱大学进行的调查中，会见了一组样本学生并询问了他们的等级分平均数是多少。该结果的直方图如下所示。应当如何解释在 2 处的尖峰？(GPA 从 0 到 4，而 2 为勉强通过。)



9. 按照 1880 年人口普查和 1970 年人口普查所报告[11]，下表给出了成年人按他们的年龄的最后一位数的分布。你可能期望十个可能的数字中的每一个对应 10%的人，但并不是这样的情况。例如，在 1980 年，所有人中的 16.8%报告了以 0 结尾的年龄——比如 30 或 40 或 50。在 1970 年，这个百分数只有 10.6%。

(a)对这两个分布绘制直方图。

(b)在 1880 年，对于数字 0 和 5 存在明显的偏向，这如何去解释。

(c)在 1970 年，这种偏向弱多了。这如何去解释？

(d)在 1880 年，偶数普遍一些，还是奇数普遍一些？1970 年呢？

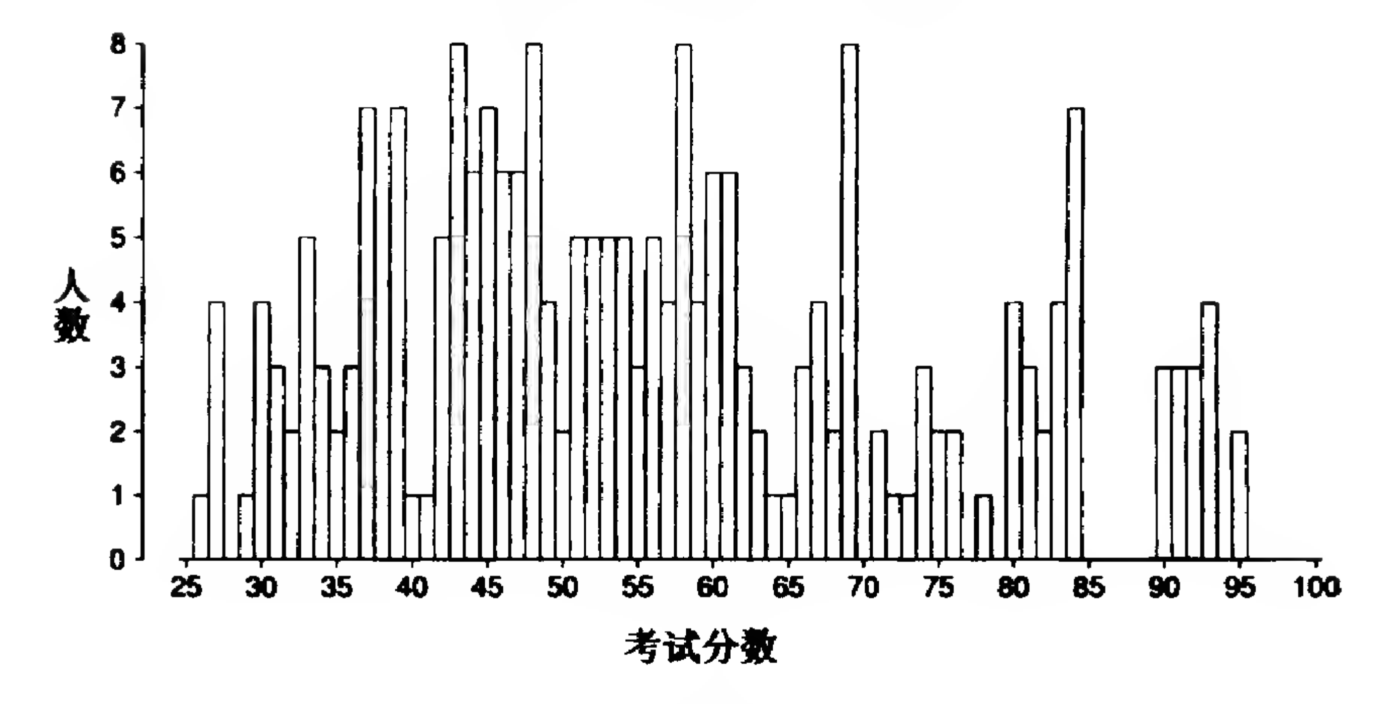
数字	1880	1970
0	16.8	10.6
1	6.7	9.9
2	9.4	10.0
3	8.6	9.6
4	8.8	9.8
5	13.4	10.0
6	9.4	9.9
7	8.5	10.2
8	10.2	10.0
9	8.2	10.1

来源：美国人口调查

10. 在芝加哥环境卫生管区，操作工程师是在有竞争性的公务员考试基础上被雇用的。1966 年，有 223 人申请 15 份工作。考试在三月十二日进行；考试分数按增加的顺序排列在下面。直方图中的每一条的高度表示得到相应分数的人数。在这一证据的基础上，考试者被指控操纵考试。[12] 为什么？

26 27 27 27 27　29 30 30 30 30　31 31 31 32 32
33 33 33 33 33　34 34 34 35 35　36 36 36 37 37
37 37 37 37 37　39 39 39 39 39　39 39 40 41 42
42 42 42 42 43　43 43 43 43 43　43 43 44 44 44
44 44 44 45 45　45 45 45 45 45　46 46 46 46 46
46 47 47 47 47　47 47 48 48 48　48 48 48 48 48
49 49 49 49 50　50 51 51 51 51　51 52 52 52 52
52 53 53 53 53　53 54 54 54 54　54 55 55 55 56
56 56 56 56 57　57 57 57 58 58　58 58 58 58 58

58 59 59 59 59　60 60 60 60 60　60 61 61 61 61
61 61 62 62 62　63 63 64 65 66　66 66 67 67 67
67 68 68 69 69　69 69 69 69 69　69 71 71 72 73
74 74 74 75 75　76 76 78 80 80　80 80 81 81 81
82 82 83 83 83　83 84 84 84 84　84 84 84 90 90
90 91 91 91 92　92 92 93 93 93　93 95 95

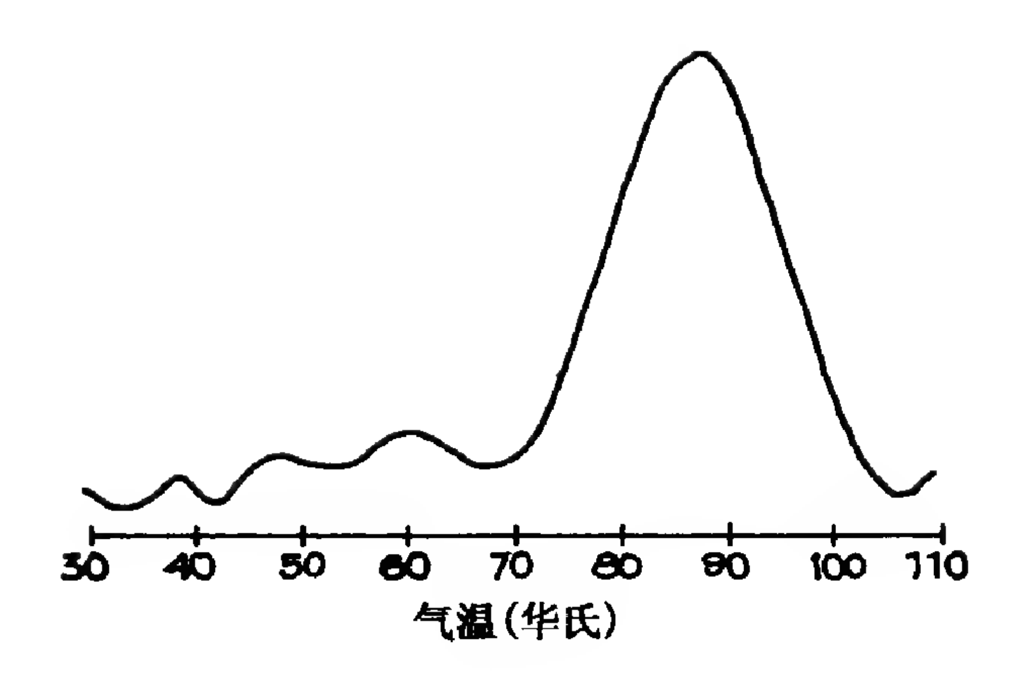


11. 六十年代后期与七十年代初期是美国骚动混乱的年份。心理学家认为骚乱(在其它事情中间)与气温有关,较热的天气使人们更加放肆[13]。但是,两位调查研究人员争辩"骚乱的频率将随着气温而增加直到 85 度,然后超过这个水平后随着气温的增加而急剧地下降。"为了支持这个理论,他们收集了 1967—1971 年期间有关 102 次骚乱的资料,包括骚乱发生城市的气温。他们按气温画了一个关于骚乱分布的直方图(草图在下页给出)。在 85 度附近有一个明确的高峰。较高的气温阻止了骚乱吗?回答是或否,并简短地解释。

9. 小结

1. 直方图以面积表示百分数。它由一组块形组成。每一个块形的面积表示在相应的小组区间中事例的百分数。

2. 采用密度尺度,每一个块形的高度等于相应小组区间中事



例的百分数除以该区间的长度。

3. 采用密度尺度，面积呈现为百分数，总面积是100%。直方图下两个数值之间的面积给出了落在那个区间内的事例的百分数。

4. 变量是研究中对象的特征。它可以是定性的也可以是定量的。一个定量变量要么是离散的要么是连续的。

5. 混杂因素有时候用交叉列表加以控制。

4
平均数和标准差

很难理解为什么统计学家通常限制自己的调查于平均数，而不着迷于更广泛的考虑。对于变化的魅力，他们的灵魂看来如同平坦的英格兰国家之一的当地人的一样迟钝，那些当地人关于瑞士的回顾是，如果可以将它的山脉扔进它的湖泊，那么两种讨厌的东西将立即去除。

——Francis Galton 爵士（英格兰，1822—1911）[1]

1. 引言

直方图可以用来概括大量数据。经常地，仅仅给出直方图的中心和关于中心的散布就有可能使得概括更为精练。（这里“中心”与“散布”是普通单词，没有任何特定的技术性意义。）图 1 中草绘了两个直方图；标出了它们的中心与散布。这两个直方图有着相同的中心，但第 2 个更分散一些——有较多的面积远离中心。就统计工作而言，必须作出确切的定义，对此存在着若干种途径。平均数通常被用来寻求中心，中位数也同样如此[2]。标准差度量了关于平均

数的散布程度，四分位数间距，则是散布的另一种测度。

图 1　中心与散布。两个直方图的中心是相同的，但第 2 个直方图更分散一些。

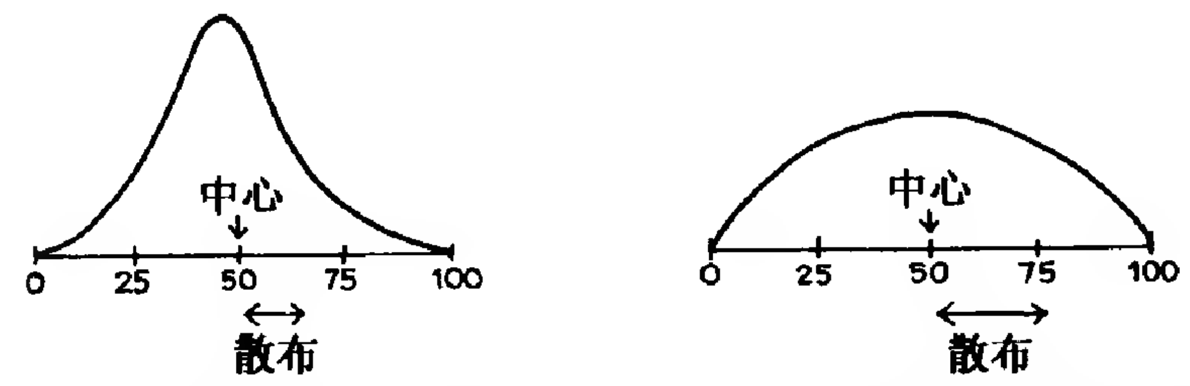
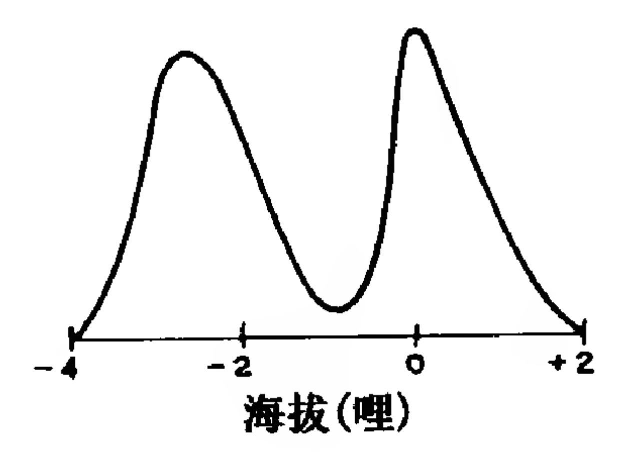


图 1 中的直方图可以通过中心和散布来概括。然而事情并非总是那么理想。譬如，图 2 给出了地球表面海拔的分布。海拔以高出海平面（+）或低于海平面（−）的哩数沿水平轴表示。在直方图下方两个海拔之间的面积给出了这两个海拔之间地球表面的百分数。该直方图有明显的两个高峰。大部分表面积由海平面以下 3 哩左右的海底所占据；或者由与海平面差不多上下的大陆平原所占据。只报告这个直方图的中心与散布将失去两个高峰[③]。

图 2　以海平面上（+）或海平面下（−）的海拔所表示的地球表面积的分布。



2. 平均数

本节的目的是考察平均数；同时也将讨论横剖面调查与纵向调查之间的差异。前后上下相联的是 HANES——1976−1980 年期间的健康与营养调查研究，在该项研究中公共卫生总署调查了年龄从 1 至 74 岁的 20 322 名美国人这样一个有代表性的剖面。目的是得到有关的基本数据——

· 诸如年龄、教育和收入这样的人口统计变量；

· 诸如身高、体重、血压和血清胆固醇含量这样的生理变量；

· 日常饮食习惯；

· 血液中铅与农药的成份；

· 疾病的流行。

随后的分析集中於这些变量之间的内在关系，这对卫生政策具有重要的影响。例如，HANES 资料揭示了在整个调查期间内血液中铅的含量下降了 37%。公共卫生总署鉴定出原因——增多了使用无铅汽油。随之出台了关于禁止添加铅的命令④。这里的关键之点在于只不过迅速浏览一下样本，再考虑了平均数的想法。

一个数列的平均数等于它们的和除以它们所含个数。

例如，数列 9,1,2,2,0 共有 5 项，首项为 9，其平均数是

$$\frac{9+1+2+2+0}{5}=\frac{14}{5}=2.8$$

样本（年龄 18—74）中的男女们看起来象什么？

· 男人平均身高 5 英尺 9 英寸，平均体重 171 磅。

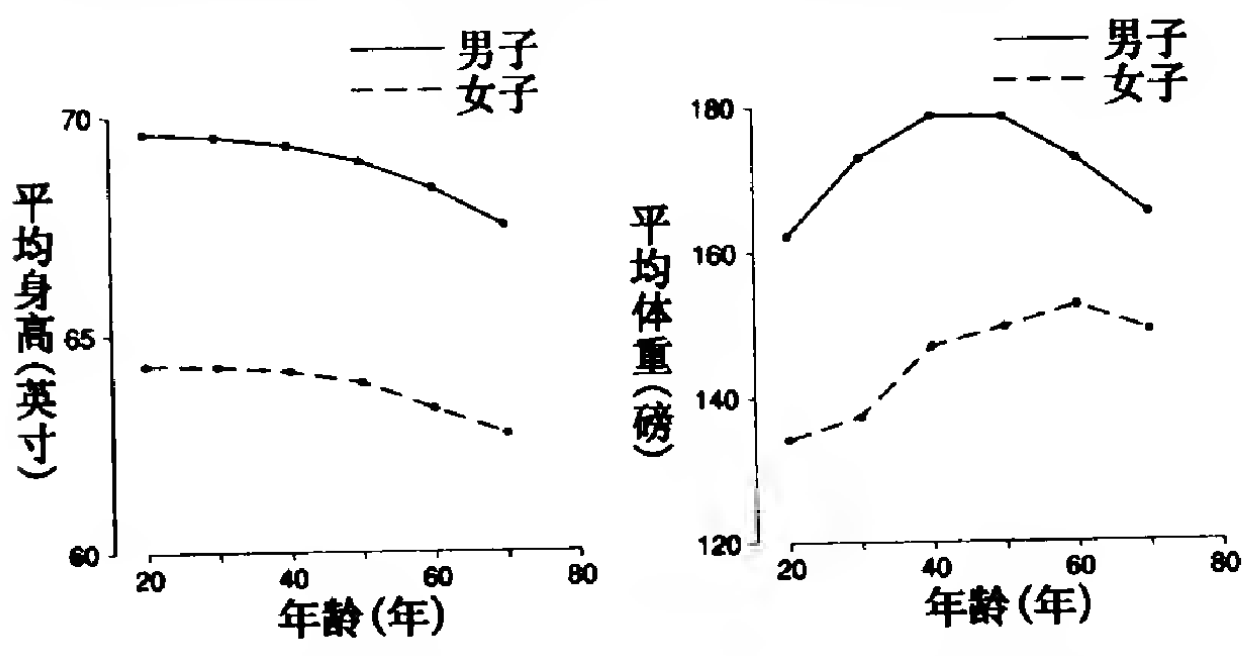
· 妇女平均身高 5 英尺 3.5 英寸，平均体重 146 磅。

他们相当丰满！

身高和体重如何与年龄相关联呢？图 3 分别地显示了由公共卫生总署所研究的各个不同年龄组中男子与妇女的平均身高和体重；这些平均数在图中用直线连接起来。平均数是概括数据的一个强有力的方法——许多直方图被浓缩入这 4 条曲线之中。可是这种浓缩仅仅通过消除个体之间的差异来达到。例如，18—24 岁男子的平均身高是 5 英尺 10 英寸。可是他们中间的 10%高过 6 英尺 1 英寸；另外的 10%则比 5 英尺 6 英寸矮。这些差异被平均数掩盖掉了。

图 3 中，男子的平均身高看来好象在 20 岁以后减少，50 年内下降大约 2 英寸。妇女也存在类似的现象。你是否该得出结论认为平均来说每个人以这样的速度变矮？不完全这样；HANES 是横

图 3 HANES 样本中 18—74 岁的男子和妇女特定年龄时的平均身高与体重。左边一张表示身高，右边一张表示体重。



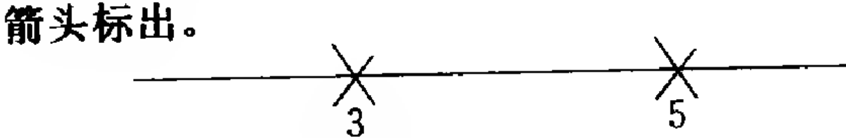
来源：为政治与社会研究的 Inter—University Consortium 提供的资料磁带。

剖面调查，不是纵向调查。在横剖面调查研究中，不同的对象在同一个时间点相互比较。在纵向调查研究中，对象在整个研究期间受到跟踪，并在不同的时间点进行自身比较。

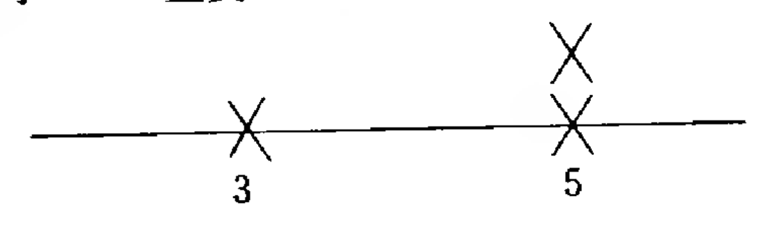
例如，图 3 中 18—24 岁的人完全不同于 65—74 岁的人。第一组大约出生于 1955 年，第 2 组大约出生于 1905 年。有迹象暗示，在这段时间内，美国人正在变得较高。这称之为在身高方面的长期趋势。在图 3 中它与年龄的效应混杂在一起，身高下落 2 英寸中的大部分似乎应归因于长期趋势。65—74 岁的人在那些年龄为 18—24 岁的人之前 50 年出生，且为此原因他们矮了 1 至 2 英寸[⑤]。

习题 A

1. (a)在下面的水平线上数 3 与数 5 用叉号标出，求这两个数的平均数并用箭头标出。



(b)对数列 3，5，5 重复(a)的做法。



(c)在水平轴上两个数用叉号标出如下。画一个箭头以指出它们的平均数。



2. 某数列有10项。每一项要么是1要么是2或者是3。如果其平均数是1,这数列必须是什么样? 如果平均数是3,这数列又是什么样? 平均数可能为4吗?

3. 下面两个数列中的哪一个有较大的平均数? 或者它们的平均数是相同的? 试在不要做任何算术运算下回答。

   (i)10,7,8,3,5,9

   (ii)10,7,8,3,5,9,11

4. (a)对下面每一个数列画出直方图(第一个数列的直方图如P69上图5所示);求出平均数;在水平轴上画一箭头指出平均数。

   (i) 1,2,2,3 (ii) 2,3,3,4 (iii) 2,4,4,6 (iv) −1,−2,−2,−3

   (b)每个数列都与数列(i)有关。例如,数列(ii)可以从数列(i)通过每一项加上1得到。数列(iii)与(iv)如何与数列(i)有关? 这些关系如何转到直方图上? 转到平均数上?

5. 某房间中10个人平均身高5英尺6英寸。身高为6英尺5英寸的第11个人进入房间。求所有11个人的平均身高。

6. 某房间中21个人平均身高5英尺6英寸。身高为6英尺5英寸的第22个人进入房间。求所有22个人的平均身高。与习题5作比较。

7. 某房间中21个人平均身高5尺6英寸。第22个人进入房间。他必须有多高才能使平均身高上升1英寸?

8. 一教员出了一份含有3个问题的测验卷,每个问题1分;班级中30%得3分,50%得2分,10%得1分,还有10%得0分。

   (a)如果班级中有10人,平均分是多少?

   (b)如果班级中有20人,平均分是多少?

   (c)不告诉你班级中有多少个人,你能算出平均得分吗?

9. 平均每小时收入是每个月由劳工统计局从商业公司提供的工资单计算而得。该局计算付出(给非管理人员)的总工资,再除以工作的总小时数。在衰退时期,平均每小时收入典型地上升。当衰退结束时,平均每小时收入常常开始下降。这可能是怎么回事?

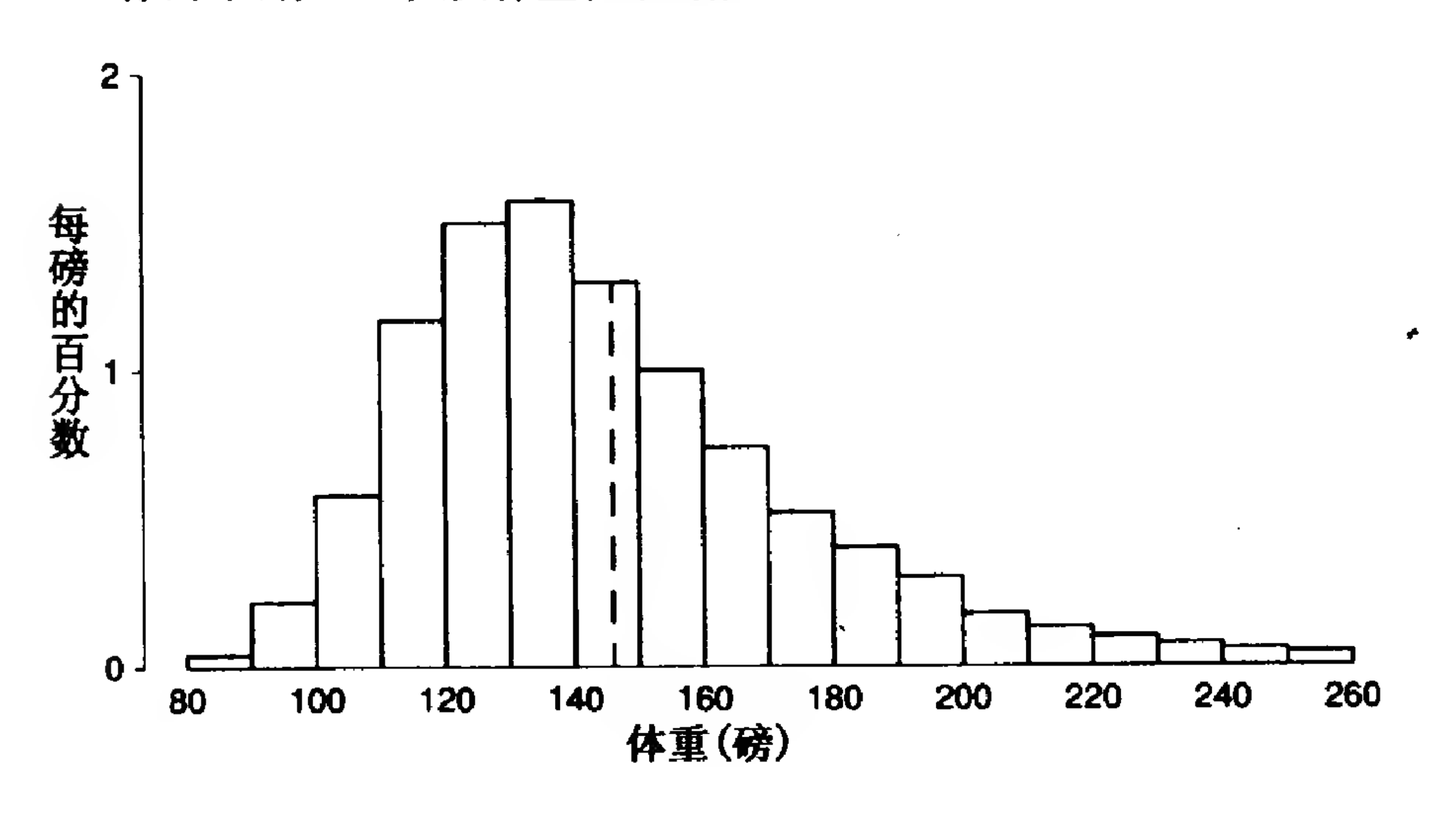
10. 图 2(64 页)中,洛矶山脉描绘在轴的左端附近,中间,还是在右端附近? Kansas(堪萨斯)州又怎么样?象 Mindnnao(明达瑙)深渊这样的在海底的地沟又怎么样?

这些习题的答案在第 675—676 页上。

3. 平均数和直方图

本节将指出平均数和中位数如何与直方图相关联。以一个例子开始,图 4 给出了 HANES 样本中 18—74 岁的 6 588 名妇女的体重直方图。平均数(146 磅)用一根垂直线标出。自然地猜测妇女中 50%其体重超过平均数,而 50%低于平均数。然而,这个猜测有点偏离。事实上,只有 41%超过平均数;余下的 59%在平均数以下。在其它场合,百分数甚至可能更远离 50%。

图 4 HANES 样本中 6 588 名妇女的体重直方图。平均数用虚线标出,只有 41%妇女体重在平均数之上。

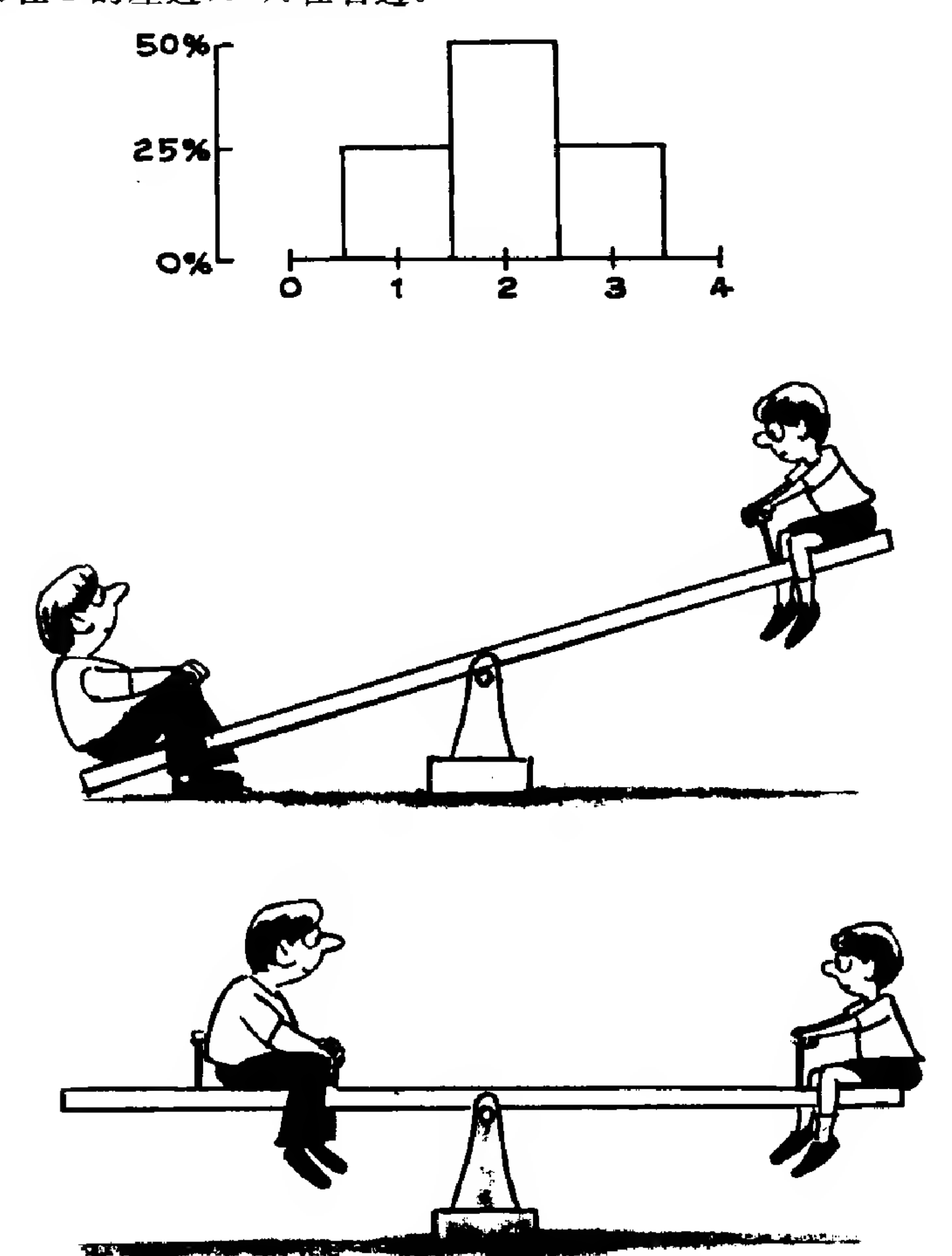


来源:由政治与社会研讨校际联合会提供的磁带。

怎么可能呢?为查明起见,最容易的是从某些虚构的数据出发——数列 1,2,2,3。这个数列的直方图(图 5)关于 2 对称。就是说,设想画一根垂直线通过数 2,直方图关于此线对折:这两个一半部分将吻合。因而平均数等于 2。如果直方图关于某个数值对称,那

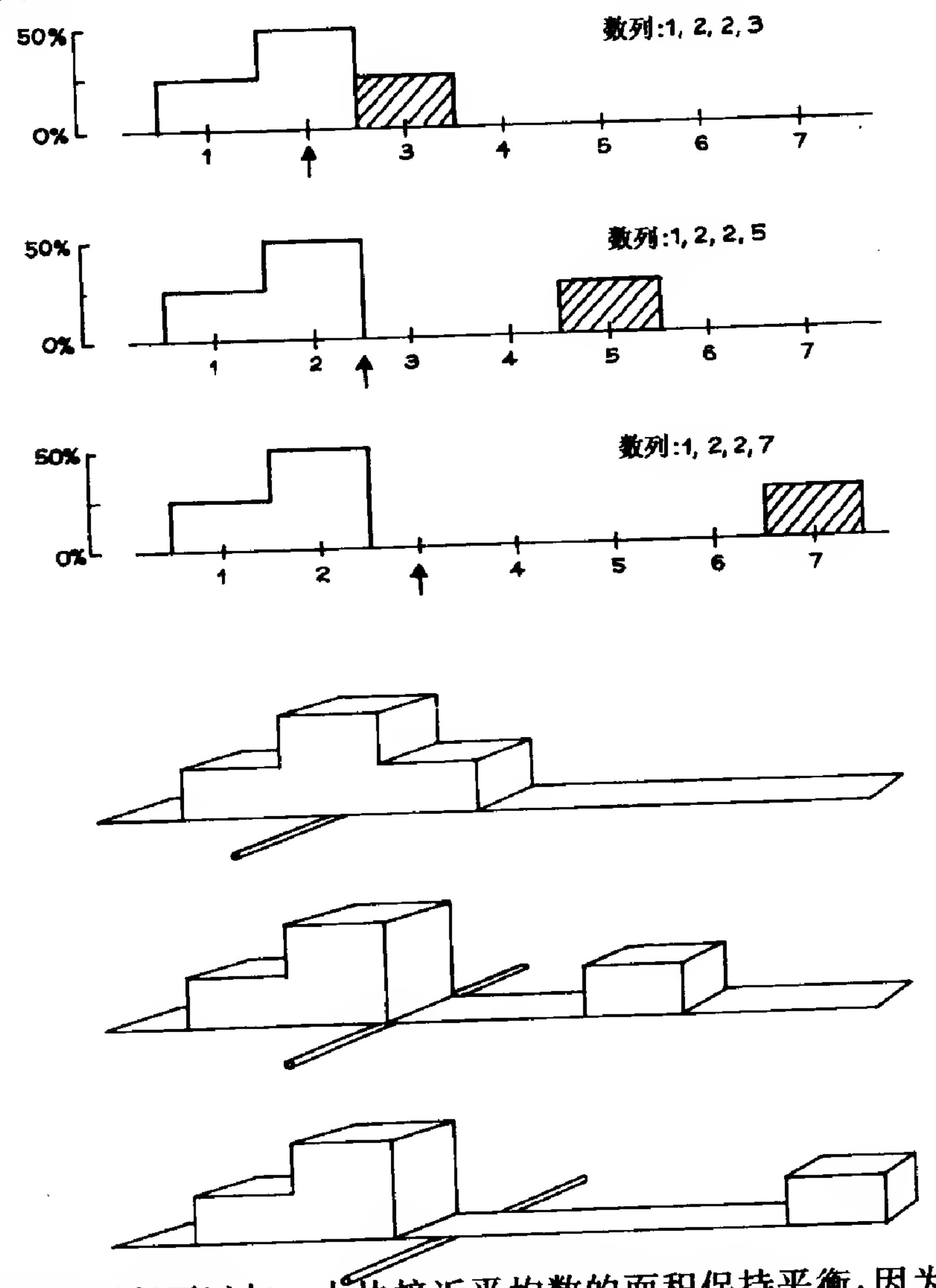
么该值等于平均数;进而,直方图以下面积的一半位于该值的左边,另一半位于该值右边。

图5 数列1,2,2,3的直方图,直方图关于平均数2对称;面积的50%在2的左边,50%在右边。



当数列1,2,2,3中的数值3增加,譬如说增加到5或7时将会发生什么?如图6所示,该数值上方的矩形移向右边,破坏了对称性。每个直方图的平均数用箭头标出,那么这个箭头随着矩形移向右边。为找出原因,假想直方图由安放在一块硬挺的无重量的木板上的木块所组成。如图6的下半部所显示的那样,用一根挺直的金属线穿过直方图。在平均数处直方图将保持平衡[6]。一小块远离

图 6 平均数 上半部给出了 3 个直方图；平均数用箭头标出。当阴影匣子移向右方时，它牵着平均数跟着走。平均数左边的面积达到 75％。下半部展示了同样的 3 个直方图，它们由木块组成并安放在一块硬挺的无重量的木板上。在平均数处支撑时直方图保持平衡。



平均数的面积可以与一大块接近平均数的面积保持平衡，因为这些面积以它们离开平衡点的距离而被加权。

在平均数处支撑时直方图保持平衡。

恰如一个跷跷板:为了与一个坐在接近中心位置的大孩保持平衡,一个小孩坐到了远离中心的位置。那就是为什么在平均数的各一边事例的百分比可以不同于 50%。

直方图的中位数是这样一个数值,在它的左右两边各具有一半面积。在图 6 的第 3 个直方图中,中位数是 2。中位数右边的面积与左边的面积相比离得远一些。因而,如果你试图在中位数处平衡直方图,它将向右边倾斜。更一般地,如图 7 所示,每当直方图有一个长长的右尾部,则平均数位于中位数的右边。体重直方图(P68 上的图 4)有一个长长的右尾部;那就是为什么平均数大于中位数的理由。

图 7 直方图的尾部。

长右尾部　　对称　　长左尾部

平均数大于中位数　　平均数与中位数相同　　平均数小于中位数

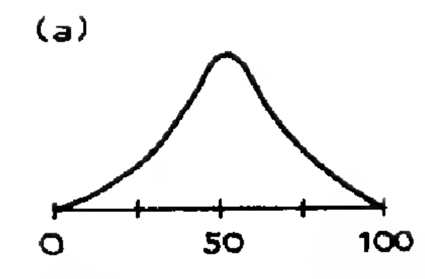
举另一个例子,1987 年美国家庭收入中位数大约为 30 800 美元。收入直方图有一个长的右尾部,且平均数较高一些——37 000 美元[7]。在处理长尾的分布时统计学家常常使用中位数而不用平均数,理由在于在某些情况平均数过多地注意了分布的极端尾部的小百分比的事例。

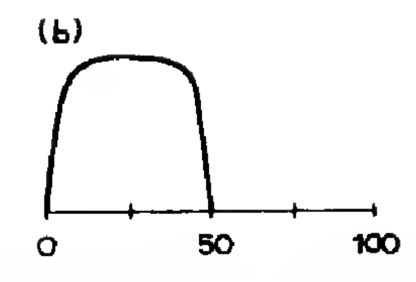
习题 B

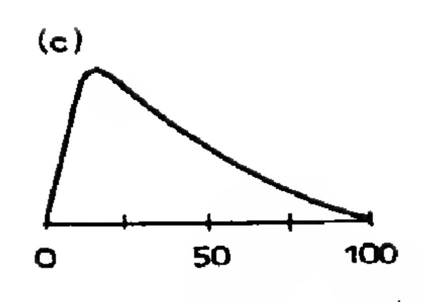
1. 下面是 3 个数列的直方图的草图。每个数列的平均数大约是________。对每一个数列,使用下面选择之一填入空白。

25　40　50　60　75

2. 习题 1 中的每一个直方图,中位数等于平均数吗?或者是在平均数的左边?还是在右边?







3. 回头观察第 47 页上的抽烟直方图。中位数大约是________。使用下面选择之一填入空白。

10　20　30　40

4. 对于这个抽烟直方图，平均数大约是 15，20，或 25？

5. 对于美国大学中的注册学生，哪一个较大一些：平均年龄还是中位数年龄。

6. 对下面每一个数列，不需要进行算术运算，指出它们按规模是否大体上围绕着 1，5，或者 10。

(a)1.3，0.9，1.2，0.8　　(b)13，9，12，8

(c)7，3，6，4　　(d)7，－3，－6，4

这些习题的答案在第 676 页上。

技术性注。数列的中位数定义为，使得一半或一半以上的项在中位数或较大一些，一半或一半以上的项在中位数或较小一些。将对 4 个数列给以阐述——

(a)1，5，7

(b)1，2，5，7

(c)1，2，2，7，8

(d)8，－3，5，0，1，4，－1

数列(a)的中位数是 5：　3 项中有 2 项为 5 或 5 以上，2 项为 5 或 5 以下。对于数列(b)，2 与 5 之间的任意数都是中位数；如果一定要选择一个的话，大多数统计学家将选择 3.5(2 与 5 之间的中点)作为“指定的”中位数。数列(c)的中位数是 2：5 项中有 4 项是 2 或 2 以上，而 3 项是 2 或 2 以下。为求数列(d)的中位数，以递增次序加以排列：

－3，－1，0，1，4，5，8

在这个数列中有7 项：4 项为 1 或1 以上，4 项为 1 或 1 以下。因此 1 是中位数。

4. 均方根

本章的下一个主要话题是标准差，它是用来度量散布的。本节

提出一个数学预备，以下面数列给予阐述

$$0,\quad 5,\quad -8,\quad 7,\quad -3$$

总体来说，这 5 个数有多大？其平均数是 0.2，但是这是大小的一个蹩脚测度：它仅仅意味着在很大程度上正数与负数相抵消。绕过这一点最简单的办法是摒弃原来的符号然后再取平均值。然而，统计学家还做了一些事：他们对数列应用了均方根的运算。词汇"均方根"(Root-mean-square)说出了如何去做这个算术运算，只要你记住了反过来阅读它的英语原文：

- SQUARE(平方)所有的项，消去符号。
- 对这些平方项取平均数(MEAN)。
- 对该平均数取平方根(ROOT)。

这些做法可以表示为一个公式，将均方根缩写为 r.m.s.

$$\boxed{\text{数列的 r.m.s.} = \sqrt{(\text{项})^2\ \text{的平均数}}\text{。}}$$

例 1. 对数列 0,5,−8,7,−3，求其平均数，和略去符号后的平均数，以及 r.m.s。

解

$$\text{平均数} = \frac{0+5-8+7-3}{5} = 0.2$$

$$\text{略去符号后的平均数} = \frac{0+5+8+7+3}{5} = 4.6$$

$$\text{r.m.s.} = \sqrt{\frac{0^2+5^2+(-8)^2+7^2+(-3)^2}{5}} = \sqrt{29.4} \approx 5.4$$

r.m.s. 稍微比略去符号后的平均数大一些。总是得出类似的结果(除非所有的项均相同这样的平凡情况。) 由于插入了取平均数这个步骤，开根号与平方不会相互抵消。

从 5.4 和 4.6 之间选择一个作为本例中数列的综合大小看来似乎并没什么两样。统计学家则使用 r.m.s.，因为它与他们必须进行的代数运算比较适合[8]。不必怀疑这种解释是否具有吸引力。每一个人起先嫌恶 r.m.s.，然而很快地就习惯用它了。

习题 C

1. (a)求数列 1,−3,5,−6,3 的平均数与 r.m.s. 大小。

   (b)对数列−11,8,−9,−3,15 做同样的运算。

2. 猜测下面数列中每一个的 r.m.s. 大约是 1,10,还是 20。不需要进行算术运算。

   (a)1,5,−7,8,−10,9,−6,5,12,−17

   (b)22,−18,−33,7,31,−12,1,24,−6,−16

   (c)1,2,0,0,−1,0,0,−3,0,1

3. 下面 3 个数列中哪一个的 r.m.s. 最小?哪一个的最大?不需要进行算术运算。

   (i)−6,−3,−1,0,5,5

   (ii)−3,−2,0,1,1,3

   (iii)−4,−4,0,2,3,3

4. (a)求数列 7,7,7,7 的 r.m.s.。

   (b)对数列 7,−7,7,−7,重复(a)的运算。

5. 数 103,96,101,104 中的每一个数几乎是 100,但有一定量的差异,求这些差异量的 r.m.s.。

6. 数列 103,96,101,104 有一个平均数,试求之。数列中的每一个数偏离平均数若干量。求偏离量的 r.m.s.。

7. 一计算机被编制程序用来预测考试分数,将它们同实际分数相比较,求出预测误差的 r.m.s.。对打印结果粗略地看一下,发现预测误差的 r.m.s. 是 3.6,以及头 10 名学生的如下结果:

   预测分:90 90 87 80 42 70 67 60 83 94

   实际分:88 70 81 85 63 77 66 49 71 69

   有否差错?

这些习题的答案在第 676—677 页上。

5. 标准差

正如本章开头引言中所提到的,考虑一个数列关于其平均数的散布常常是有益的。这种散布通常由一个称之谓标准差,或 SD

的量来度量。SD 度量了偏离平均数的大小;它是一类平均偏差。打算要做的是解释和实际数据有关的 SD,然后看看如何去计算 SD。

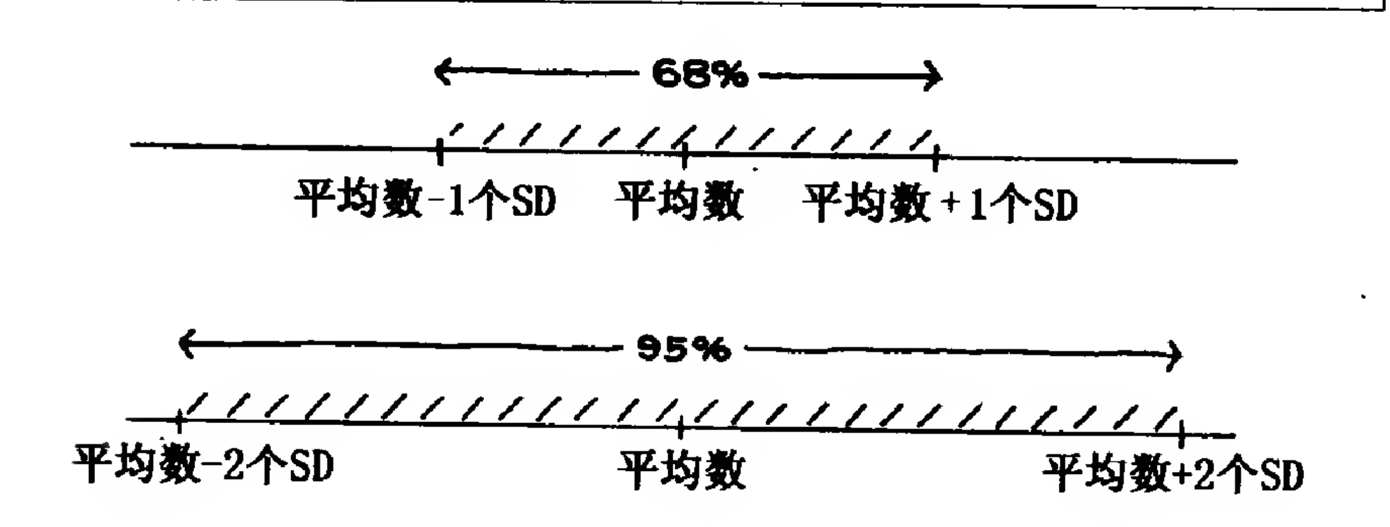
在(第 2 节)HANES 样本中有 6 588 名年龄为 18—74 岁的妇女。她们身高的平均数是 63.5 英寸,SD 为 2.5 英寸。平均数告诉我们大多数妇女为 63.5 英寸左右的身高。但是离平均数有一定偏差。有些妇女比平均数高一些,而有些则矮些。这些偏差有多大呢?那就是 SD 起作用之处。

SD 指出了数列中的数离它们的平均数有多远。数列中大多数项离开平均数大约 1 个 SD 左右。极少数项将离开 2 个或 3 个 SD 以上。

2.5 英寸的 SD 意指在 HANES 中的许多妇女与平均身高约相差 1 或 2 或 3 英寸:1 英寸小于半个 SD,3 英吋则在 1 SD 与 2 SD 之间。很少量的妇女与平均身高差 5 英吋(2 SD)以上。

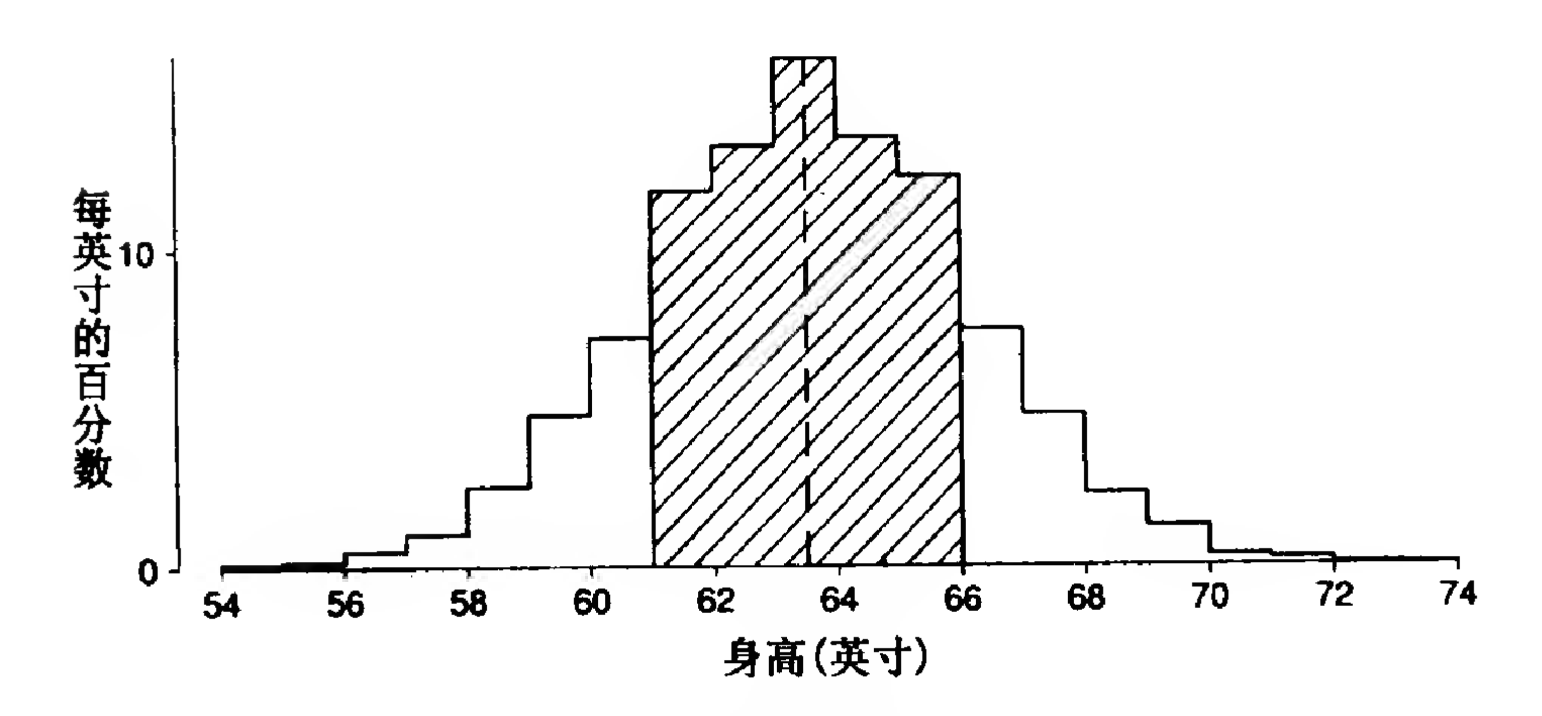
有一个约略的准则使得这个概念更数量化,它应用于许多数据组。

粗略地,数列中 68%(三分之二)的项在离平均数的 1 SD 范围内,其余的 32%离得较远。粗略地,95%(20 分之 19)的项在距平均数的 2 个 SD 范围内,其余的 5%则远离之。这对许多数列是如此,但不是所有的数列都这样。



就 HANES 数据而言，67%妇女距平均身高相差约 1 个 SD 或不到 1 个 SD，94%妇女距平均数相差 2 个 SD 或不到 2 个 SD。样本中只有 1 名妇女其身高距平均数相差多于 4 个 SD。对于这组数据，约略准则十分成功。直方图如图 8 所示。平均数用一根垂直线标出，在平均数 1 SD 以内的范围绘以阴影。

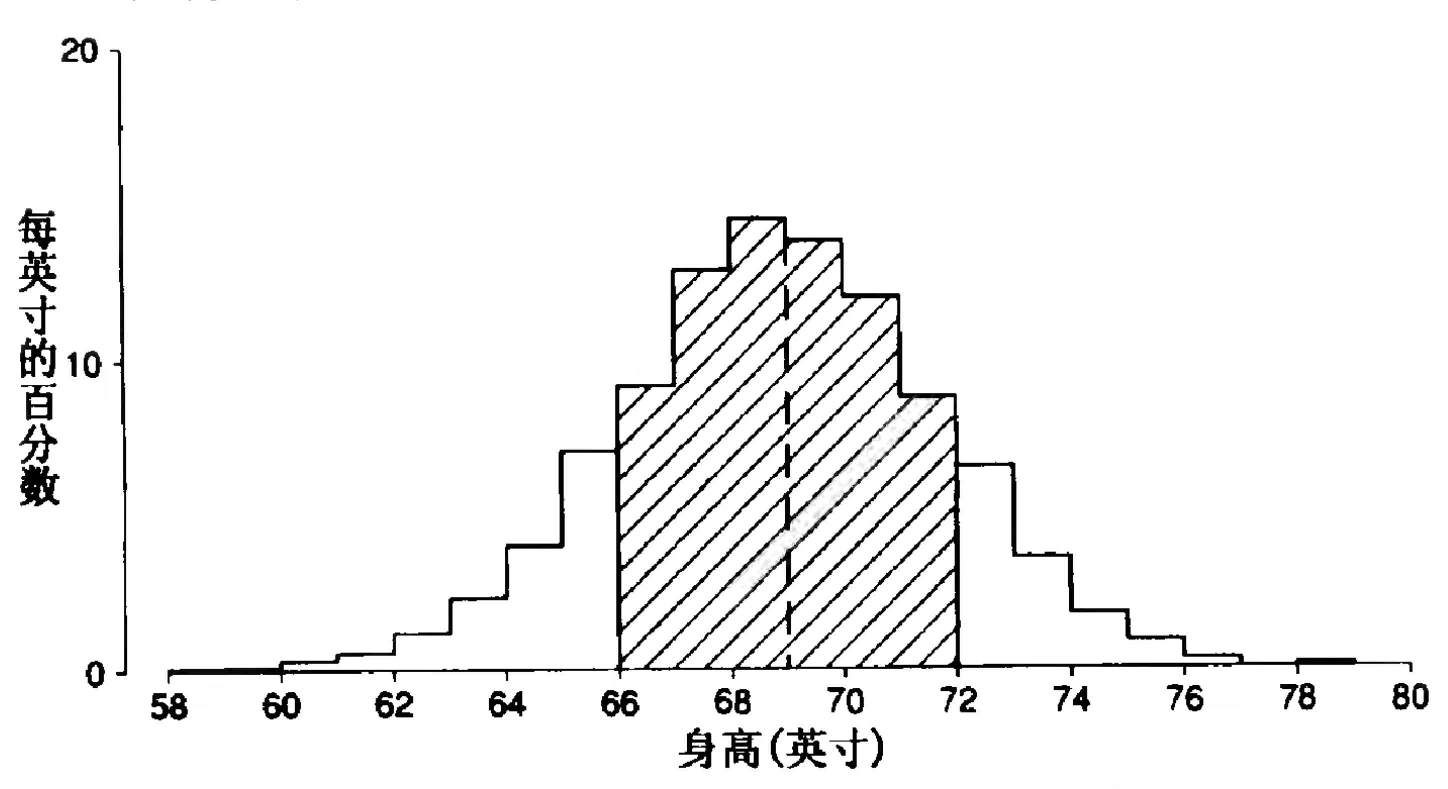
图 8 HANES 样本中年龄为 18—74 岁的 6 588 名妇女的身高直方图与 SD。63.5 英寸的平均数用一根垂直虚线标出。在平均数 1 个 SD 内的范围用阴影表示：67%妇女与平均数相差 1 个 SD(2.5 英寸)或不到 1 个 SD。



下一个例子是样本中男子的身高。身高平均数是 69 英寸；SD 是 3 英寸。图 9 给出了直方图。关于身高，71%男子与平均数相差不到 1 个 SD，96%男子与平均数相差不到 2 个 SD。没有一名男子与平均数相差超过 4 个 SD。再一次地，68—95%准则十分成功。

68%与 95%这两个数字是从哪儿来的呢？将在下一章中提及[9]。

图 9 HANES 样本中年龄为 18—74 岁的 5 916 名男子身高直方图与 SD。69 英寸的平均数用一根垂直虚线标出。在平均数 1 个 SD 内的范围用阴影表示：71%男子与平均数相差 1 个 SD(3 英寸)或不到 1 SD。





大约三分之二的 HANES 妇女与平均数相差不到 1 个 SD

习题 D

1. 公共卫生总署发现，对 HANES 中年龄为 11 岁的男孩来说，平均身高约为 146cm，SD 约为 8cm。仅仅粗略地估计，在该项研究中有多少百分数的 11 岁男孩其身高在 138cm 与 154cm 之间？又有多少在 130 与 162cm 之间？

2. 下面表列中的每一个有平均数 50。哪一列数在平均数附近的散布程度最大？哪一列最小？

(i) 0,20,40,50,60,80,100

(ii) 0,48,49,50,51,52,100

(iii)0,1,2,50,98,99,100

3. 下面数列中的每一个有平均数 50。哪一列数在平均数附近的散布程度最大？哪一列最小？

(i) 47,49,50,51,53

(ii) 46,48,50,52,54

(iii)46,49,50,51,54

4. 下面数列中的每一个有平均数 50。对每一列，猜测其 SD 大约是 1,2，或是 10。（这不需要作任何算术运算。）

(a)49,51,49,51,49,51,49,51,49,51

(b)48,52,48,52,48,52,48,52,48,52

(c)48,51,49,52,47,52,46,51,53,51

(d)54,49,46,49,51,53,50,50,49,49

(e)60,36,31,50,48,50,54,56,62,53

5. HANES 样本中人的年龄的 SD 大约是________。利用下面的选择之一填入空格，并简略地解释。（此项调查在第 2 节中讨论过；年龄的范围是 1—74 岁。）

5 年　　　　20 年　　　　50 年

6. "教育水平"是指已完成正规学校教育的年数。HANES 样本中年龄为 25 岁及 25 岁以上的人的教育水平的 SD 大约是________。利用下面的选择之一填入空格；简略地解释。

1 年　　4 年　　9 年　　3 磅　　5mm

7. 下面是 3 个数列直方图的草图。将草图与描述相匹配。某些描述将是多余的。对每一种情况给出你的理由。

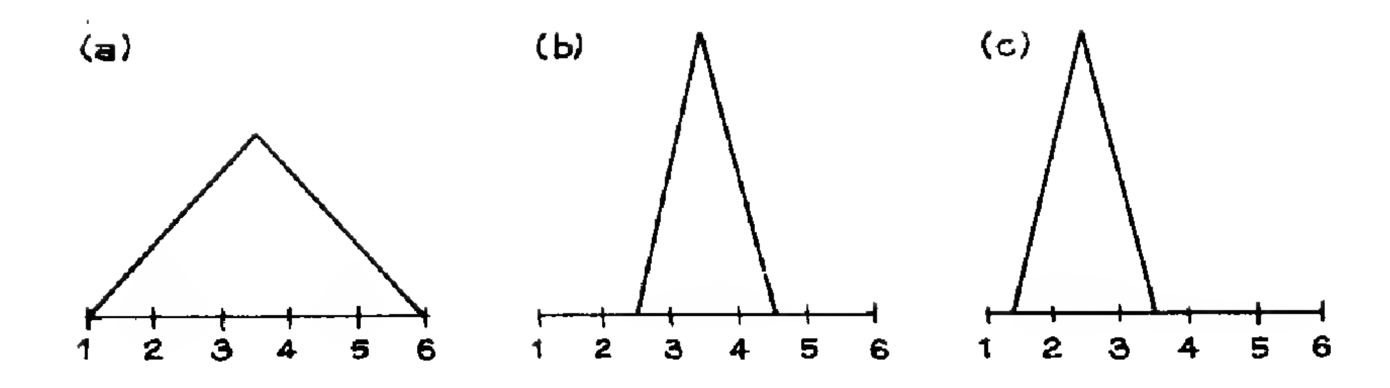
(i)平均数≈3.5,SD≈1　　(iv)平均数≈2.5,SD≈1

(ii)平均数≈3.5,SD≈0.5　　(v) 平均数≈2.5,SD≈0.5

(iii)平均数≈3.5,SD≈2　　(vi)平均数≈4.5,SD≈0.5

8. 构造 10 个数的数列使得其 SD 尽可能地大并且——

(a)每一个数要么是 1 要么是 5



(b)每一个数要么是 1 要么是 9

(c)每一个数要么是 1 要么是 5 或者 9，它们中至少有 2 个数是 5。

此习题不需要任何算术运算。

9. 重复习题 8，此时要求所得 SD 尽可能地小。

10. (假设的)在临床试验中，数据收集通常在"基线"开始，在对象被征集参加试验但是尚未随机地指派进处理组或者对照组之前的时候。数据收集继续到结束与对象保持的联系为止。两个关于预防心脏病的临床试验报告了有关体重的基线数据，如下所示。在这些试验中的某一个，随机化没搞好。是哪一个，为什么？

		人数	平均体重	SD
(i)	处理	1 012	185 磅	25 磅
	对照	997	143 磅	26 磅
(ii)	处理	995	166 磅	27 磅
	对照	1 017	163 磅	25 磅

11. 一调查研究人员在某城取 100 名 18—24 岁的男子作为样本。另一调查研究人员取 1 000 名这样的男子作为样本。

(a)哪一位调查研究人员将在他所取样本中得到男子身高的较大的平均数？或者这两个平均数大致相等？

(b)哪一位调查研究人员将在他所取得样本中得到男子身高的较大 SD？或者这两个 SD 大致相等？

(c)哪一位调查研究人员有可能得到这 1 100 名男子样本中的最高者？或者对这两位调查研究人员来说，这种机会是相等的？

(d)哪一位调查研究人员有可能得到这 1 100 名男子样本中的最矮者？或者对这两位调查研究人员来说，这种机会是相等的？

这些习题的答案在第 677—678 页上

6. 计算标准差

为求一个数列的标准差，在数列的项中一次取 1 项。每一项离平均数相差一些量，可能是 0：

与平均数的偏差=项－平均数。

SD 是离平均数的这些偏差的 r. m. s. ：

$$\boxed{\text{SD}=\sqrt{(\text{与平均数的偏差})^2\ \text{的平均数}}。}$$

例 2. 求数列 20，10，15，15 的 SD。

解 第一步是求平均数：

$$\text{平均数}=\frac{20+10+15+15}{4}=15。$$

第二步是求离平均数的偏差：即从每一项减去平均数。偏差为

5 　　 －5 　　 0 　　 0

最后一步是求这些偏差的 r. m. s. ：

$$\begin{aligned}\text{SD}&=\sqrt{\frac{5^2+(-5)^2+0^2+0^2}{4}}\\&=\sqrt{\frac{25+25+0+0}{4}}\\&=\sqrt{\frac{50}{4}}=\sqrt{12.5}\approx 3.5\end{aligned}$$

计算完毕。

SD 得出与数据相同的单位。例如，如果身高以英寸来测量，那末 SD 也得出为英寸。在计算过程当中的平方步骤将单位改变成平方英寸，但是最后步骤取平方根将答案转回到原来的单位[10]。

不要混淆数列的 SD 与 r. m. s. 。SD 是 r. m. s. ，它不是数列中原先的数的 r. m. s. ，而是它们离平均数的偏差的 r. m. s. 。

习题 E

1. 猜测下面两个数列中的哪一个具有较大的 SD。通过对这两个数列计算 SD

来检验你的猜测。

(i) 9,9,10,10,10,12

(ii)7,8,10,11,11,13

2. 某人告诉你如何计算数列 1,2,3,4,5 的 SD:

平均数是 3,因此离平均数的偏差是

$-2 \quad -1 \quad 0 \quad 1 \quad 2$

去掉符号,平均偏差为

$$\frac{2+1+0+1+2}{5}=1.2$$

那就是 SD。

这样对吗? 回答是或否,并简略地解释。

3. 某人告诉你如何计算数列 1,2,3,4,5 的 SD:

平均数是 3,因此离平均数的偏差是

$-2 \quad -1 \quad 0 \quad 1 \quad 2$

0 不计算,因此 r. m. s. 偏差为

$$\sqrt{\frac{4+1+1+4}{4}}\approx 1.6$$

那就是 SD。

这样对吗? 回答是或否,并简略地解释。

4. 三位教员正在比较他们的期终成绩;每一位有 99 名学生。在 A 班,1 名学生得 1 分,另一名得 99 分,其余的人得 50 分。在 B 班,49 名学生得 1 分,1 名学生得 50 分,49 名学生得 99 分。在 C 班,1 名学生得 1 分,1 名学生得 2 分,1 名学生得 3 分,等等,一直到 1 名学生得 99 分。

(a)哪一个班具有最大的平均数? 还是它们的平均数是相同的?

(b)哪一个班具有最大的 SD? 还是它们的 SD 是相同的?

(c)哪一个班具有最大的变化范围? 还是它们的变化范围是相同的?

5. (a)对下面每一个数列,求出平均数,离平均数的偏差,和 SD。

(i) 1,3,4,5,7

(ii)6,8,9,10,12

(b)数列(ii)与数列(i)如何相关联?这种关系如何延续到平均数?如何延续到离平均数的偏差? 如何延续到 SD?

6. 对下面两个数列重复习题 5:

(i) 1,3,4,5,7

(ii)3,9,12,15,21

7. 对下面两个数列重复习题5:

(i) 5,−4,3,−1,7

(ii)−5,4,−3,1,−7

8. (a)加利福尼亚州长提出给所有的州政府雇员一律增加每月70美元。这对于州政府雇员的平均月薪将会有何影响?对于SD呢?

(b)整个政府部门的工资5%的递增将对平均月薪有何影响?对于SD呢?

9. 数列17,17,17,17,17的r.m.s.是什么?其SD是什么?

10. 对于数列107,98,93,101,104,哪一个较小一些——是r.m.s.还是SD?不需要进行算术运算。

11. SD可能会是负数吗?

12. 对于一个正数数列,SD可能会大于平均数吗?

13. 一个数列有10个数,各数要么是1,2,要么是3。

(a)平均数是2,SD是0,该数列是什么样的?

(b)SD是1,该数列是什么样的?

(c)SD能大于1吗?

14. 一教员搞了一次总分为20的测验。评分准则是使得分数必须是5的倍数。她得到了如下分布:

得分	百分数
20	40
15	30
10	20
5	10

(a)假如班中有10人。你能得出分数的平均数与SD吗?

(b)假如班中有20人。你能得出分数的平均数与SD吗?

(c)假如你不知道班中有多少人,你能得出分数的平均数与SD吗?

15. HANES样本中男子具有69英寸的平均身高,SD是3英寸。某日,将从这些男子中随机地选出1人。你必须猜测他的身高,你该猜测什么?你希望这个猜测应该相差大约1/2英寸,1英寸,还是3英寸?

16. 在习题15中,某日将随机地选出一系列男子。在每个男子选出之后,他的真实身高将同你的猜测相比较,从而看到你的猜测相差多少。所相差量的r.m.s.应是________。

这些习题的答案在第678—679页上。

技术性注。存在另外一个计算 SD 的办法，在某些场合它更为有效[11]：

$$SD=\sqrt{(项)^2\ 的平均数-(项的平均数)^2}。$$

7. 使用统计计算器

大多数统计计算器产生的不是 SD，而是稍微大一些的数 SD^+。(SD 与 SD^+ 之间的区别将在第 26 章第 6 节中更详尽地阐述。)为查明你的计算器做的是什么运算，以数列 −1，1 输入：如果计算器给你的是 1，那么它得出的是 SD；如果计算器给你的是 (1.41…)，那么它得出的是 SD^+。如果你得到的是 SD^+ 而你想得到 SD，则必须乘上一个换算因子。它依赖于数列中的项数。数列有 10 项，换算因子是 $\sqrt{9/10}$。有 20 项，则为 $\sqrt{19/20}$。一般地，

$$SD=\sqrt{\frac{项数-1}{项数}}\times SD^+。$$

8. 复习题

复习题可能包含前面几章的内容。

1. 求数列 12，1，3，11，6，3 的平均数与 SD。

2. (a)下面两个数列具有相同的平均数 50。哪一个数列具有较小的 SD，为什么？不必进行计算。

   (i) 50，40，60，30，70，25，75

   (ii)50，40，60，30，70，25，75，50，50，50

   (b)对下面两个数列重复(a)的工作。

   (i) 50，40，60，30，70，25，75

   (ii)50，40，60，30，70，25，75，99，1

3. 有一个数列

   0.7　1.6　9.8　3.2　5.4　0.8　7.7　6.3　2.2　4.1

   8.1　6.5　3.7　0.6　6.9　9.9　8.8　3.1　5.7　9.1

   (a)不做任何算术运算，猜测平均数大约是 1，还是 5，或者是10。

   (b)不做任何算术运算，猜测 SD 大约是 1，还是 3，或者是 6。

4. 正确或者错误,并简略地解释——

(a)如果对一个数列的每一项加 7,那末平均数加 7。

(b)如果对一个数列的每一项加 7,那末 SD 加 7。

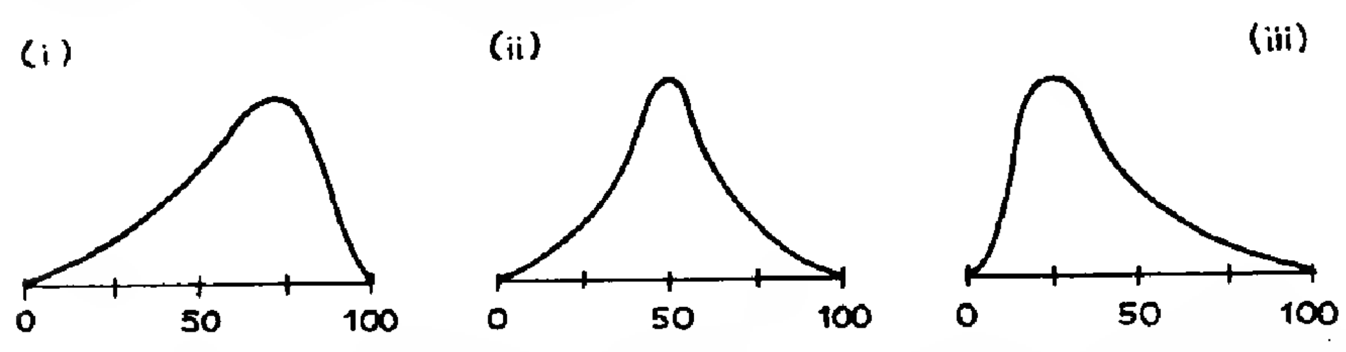
(c)如果对数列的每一项加倍,那末平均数加培。

(d)如果对数列的每一项加倍,那末 SD 加倍。

(e)如果对数列的每一项改变符号,那末改变了平均数的符号。

(f)如果对数列的每一项改变符号,那末改变了 SD 的符号。

5. 下面是 3 个数列直方图的草图。



(a)以搅乱了的次序,平均数是 40,50,60。将直方图与平均数相匹配。

(b)直方图(iii)的 SD 大约是 5,15,或者是 50?

(c)正确还是错误,并给予解释:直方图(i)的 SD 比起直方图(iii)来非常地小。

6. 一项关于大学生的研究发现男生具有大约 66 公斤的平均体重和大约 9 公斤的 SD。女生具有大约 55 公斤的平均体重和 9 公斤的 S D。

(a)以磅为单位求平均数与 SD(1 公斤=2.2 磅)。

(b)仅仅是粗略地,有多少百分数的男生其体重在 57 公斤与 75 公斤之间?

(c)如果将男女生放在一起,他们体重的 SD 将会小于 9 公斤,恰好是 9 公斤,或者大于 9 公斤? 为什么?

7. 在 HANES 样本中,男孩的平均身高在 9 岁时为 136cm,在 11 岁时为 146cm。在 11 岁时,所有孩子的平均身高是 147cm[12]。

(a)平均来说,11 岁的男孩较高于女孩吗?

(b)估计 10 岁男孩的平均身高。

8. 在美国，年龄为 25 岁及以上的人，其收入的平均数还是中位数将比较高些？已完成的正规教育年数呢？

9. 某项研究是关于学院新生的入学年龄。其 SD 大约是 1 个月，或者 1 年，或者是 5 年？为什么？

10. 对于 HANES 样本中的男子，不同年龄组的平均收入随着年龄而增加直到 50 岁左右，然后逐渐递减。正确还是错误并给予解释：资料显示了典型男子的收入随着年龄而增加直到 50 岁左右，然后开始减少。

11. 某法律学校的新生具有 LSAT（法律学校能力考试）平均分 33 及其 SD 为 5。某日，这些学生中的 1 名将被随机地挑选出来。现在你必须猜测他的分数；猜测分将与真实分相比较，看看相差有多远。每相差 1 分将付出 1 美元的代价。（例如，如果猜测是 38 分而实际是 31 分，你将不得不付出 7 美元。）

(a)最佳猜测是 30，33，还是 36？

(b)你可能的损失大约该是多少：1 美元，5 美元，或者10 美元？

12. 如同习题 11，但是有一整系列学生被选择。你的损失的 r. m. s. 将大约是________。

13. HANES 中的男子，平均收缩血压是 121mm，SD 是 22mm。指出下面血压中的每一个是否非常高，或者非常低，或者在平均数附近：

80mm　115mm　135mm　210mm

14. 1985 年全国健康调查资料展示在下。例如，18—64 岁的人，70%在该年每天吃早饭，相比于 65 岁及以上的人，有 90%每天吃早饭。是或否：资料显示，当人们变老时，他们变得更富健康意识并且接受更多的有利于健康的生活方式。解释你的回答。

年龄	吃早饭	现时喝酒者	现时抽烟者
18—64	70%	40%	35%
65 及以上	90%	10%	15%

注：百分数为四舍五入。　　来源：统计摘要 1988 表 178

9. 小结

1. 一个典型的数列可以用它的平均数和标准差(SD)来概述。

2. 数列的平均数$=\frac{\text{各项之和}}{\text{项数}}$

3. 平均数确定了直方图的中心，意思是说在平均数处获得支撑时，直方图保持平衡。

4. 直方图以下的一半面积位于中位数的左边，另一半位于右边。中位数是确定直方图中心的另一个方法。

5. 数列的 r. m. s. 度量了在去除符号后数列的项有多大。

6. 数列的 r. m. s. $=\sqrt{(\text{项})^2\text{ 的平均数}}$

7. SD 度量了关于平均数的距离。数列中每一个数离开平均数一定的量。SD 是这些离散量的一种平均大小。

8. SD 是关于平均数的偏差的 r. m. s.

$$\text{SD}=\sqrt{(\text{关于平均数的偏差})^2\text{ 的平均数}}$$

9. 在一个数列中约略有 68%的项在平均数的 1 个 SD 范围内。大约 9 的数列均如此。这对许多数列是适合的，但不是对所有5%的项在平均数的 2 个 SD 范围内。

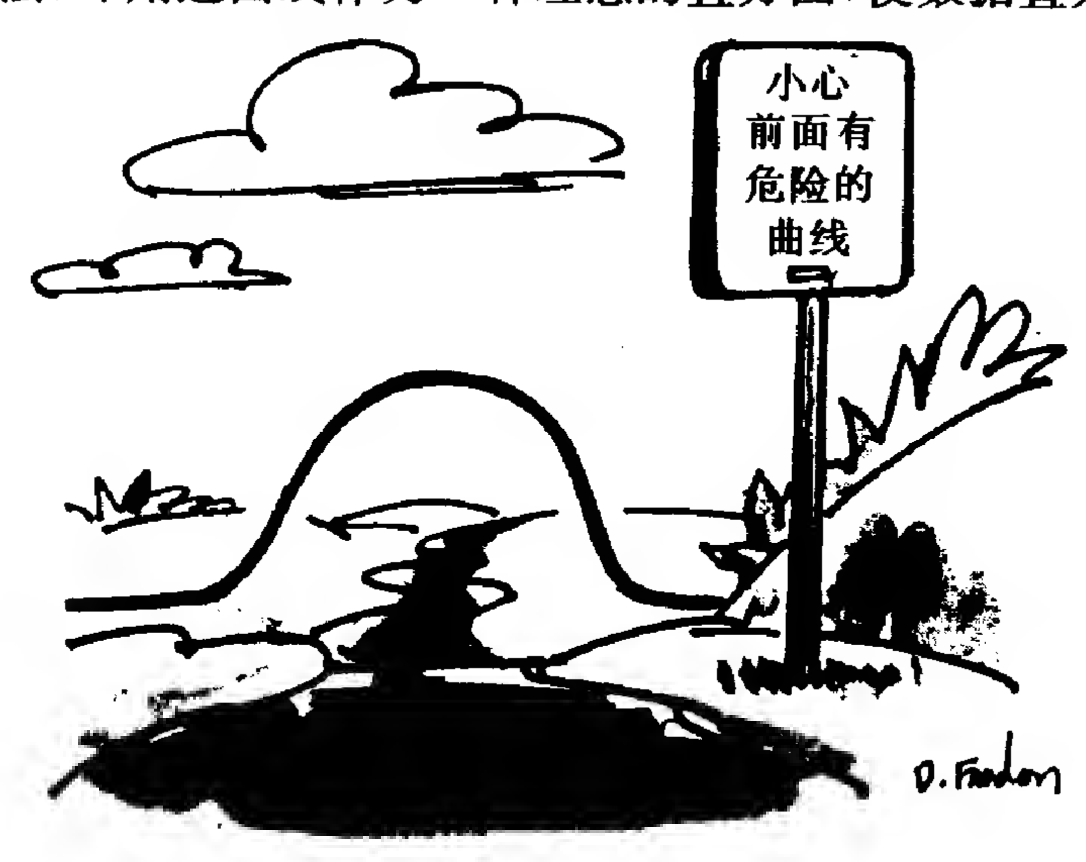


Drawing by Dana Fradon; © 1977 The New Yorker Magazine, Inc.

5
数据的正态近似

1. 正态曲线

正态曲线大约在 1720 年由 Abraham de Moivre 所发现，当时他正在发展机会数学。(他的工作将在本书第四部分与第五部分再次讨论。)大约在 1870 年，比利时数学家 Adolph Quetelet 产生了一种想法；即用这曲线作为一种理想的直方图，使数据直方图可以



与之比较。

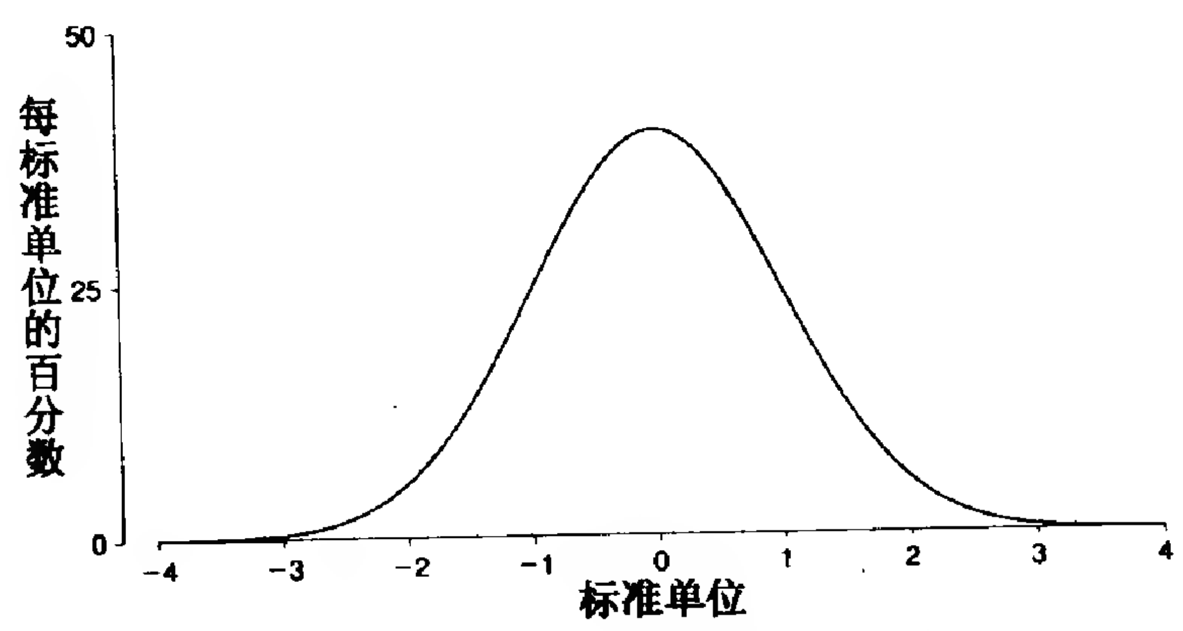
正态曲线有一个看起来令人可畏的公式：

$$y=\frac{100\%}{\sqrt{2\pi}}e^{-x^2/2},\quad 其中\ e=2.71828\cdots。$$

这个等式包含了数学史上最著名的数中的三个：$\sqrt{2}$，π，和 e。这稍微有点儿离题。你将发现即使一点公式也不用，也能容易地通过图表和正态曲线打交道。曲线的图像在图 1 中绘出。



图 1 正态曲线。



该图像的一些特征将是重要的。首先，图像关于 0 点对称。就是说，曲线在 0 点右边的部分是左边部分的镜像。其次，曲线下面的总面积等于 100%。（面积按百分数给出，因为纵轴使用了密度尺度。）最后，曲线总是在水平轴的上方；它似乎在 3 与 4 之间终止，但那仅仅是因为曲线在那里变得如此地低的缘故。只有大约 6/100000 的面积在 −4 到 4 的区间外面。

求确定值之间正态曲线下的面积将是有用的。例如，

在 −1 与 +1 之间正态曲线下的面积大约是 68%；

在 −2 与 +2 之间正态曲线下的面积大约是 95%。

求这些面积是一件查表的事（或者按功能健全的计算器上的按钮）；该表将在第 2 节中阐述。

许多有关数据的直方图只要它们是按照正态曲线的相同尺度绘制的，在形状上它们与正态曲线相象。使水平尺度相匹配涉及标

准单位[①]。一个数值通过察看它在平均数之上或平均数之下多少个SD而被转换成标准单位。在平均数之上的数值给予一个加号；在平均数以下的则得一减号。图1中的水平轴就是按标准单位的。

标准单位个数是指一个数值在平均数之上或者之下多少个SD。

以HANES样本中18—74岁的妇女为例。其中一名妇女身高68.5英寸。她的身高是多少个标准单位？所有妇女的平均身高是63.5英寸；身高的SD是2.5英寸。我们的对象比平均数高5英寸，5英寸是2个SD。换算成标准单位，她的身高是+2。

例1. 对于HANES样本中18—74岁的妇女——

(a)将下面各数换算成标准单位

(i)66英寸 (ii)58.5英寸 (iii)63.5英寸

(b)求−1.2标准单位的身高。

解(a)对于(i)，66英寸是平均数之上2.5英寸；那是平均数之上1个SD。按标准单位，66英寸是+1。对于(ii)，58.5英寸是平均数之下5英寸，即平均数之下2个SD。按标准单位，58.5英寸是−2，对于(iii)，63.5英寸是平均数，因此它与平均数相距0个SD；63.5英寸按标准单位是0。

(b)身高是平均数以下1.2SD，1.2×2.5英寸=3英寸。身高是

63.5英寸−3英寸=60.5英寸

即为答案

图2(见下一页)使用了标准单位。在该图中，将HANES样本中18—74岁的妇女身高直方图与正态曲线相比较。直方图的水平轴以英寸为单位；正态曲线的水平轴以标准单位为单位。两者如例1中所指出的那样配伍。例如，66英寸正好在+1上方，58.5英寸正好在−2上方。

图 2 妇女的身高直方图同正态曲线相比较。正态曲线的水平尺度配上关于身高的标准单位。直方图仿效正态曲线相当的好。直方图下面在 61 英寸与 66 英寸之间的面积(关于身高在离平均数 1 个 SD 范围内的妇女的百分数)大致等于在正态曲线下面 −1 与 +1 之间的面积——68%

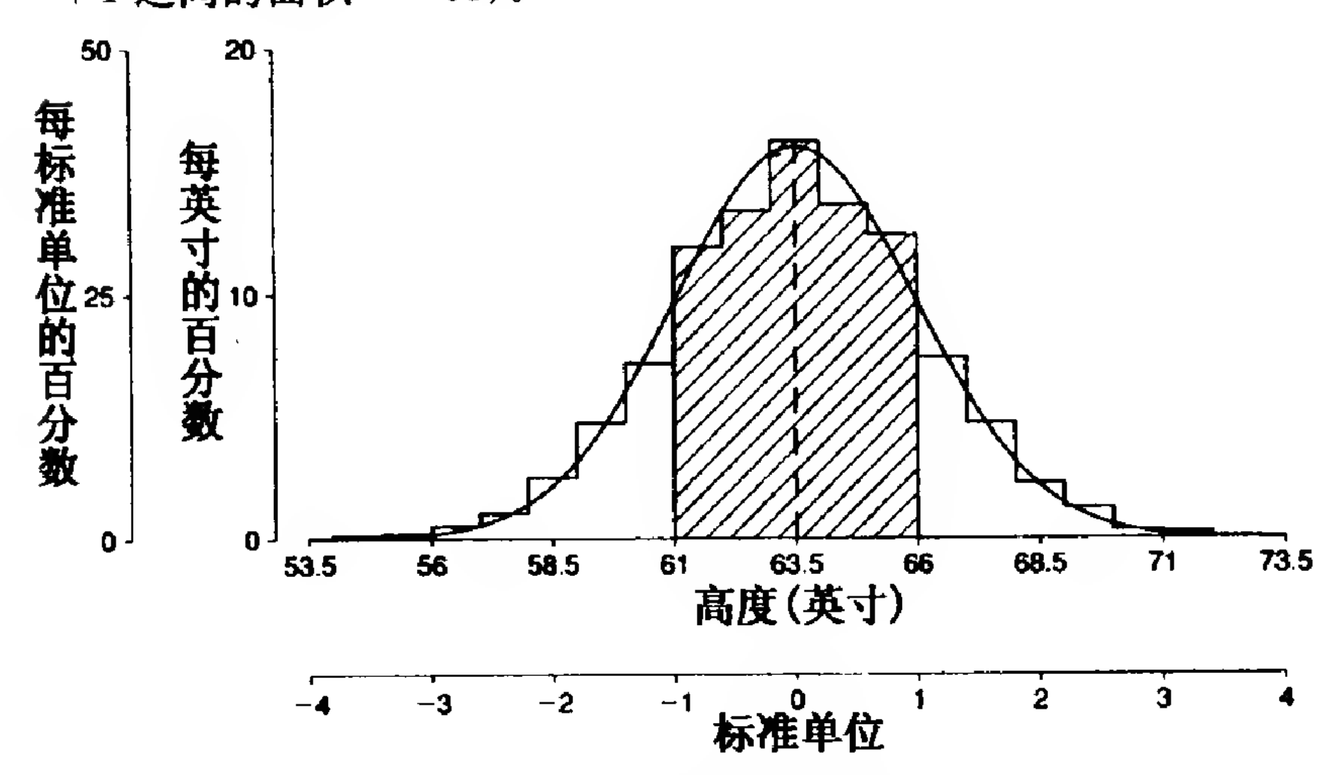


图 2 中也有两根纵轴,直方图是相对于里面的那根按每英寸的百分数而绘制的。正态曲线是相对于外面的那根按每标准单位的百分数而绘制的。为了察看这两个尺度是怎样配伍的,取最高值:每标准单位 50%与每英寸 20%相配,因为对于每个标准单位有 2.5 英寸,所以

每标准单位 50%=每 2.5 英寸 50%=每英寸 20%。

同样地,每标准单位 25%与每英寸 10%相配伍。任何其它一对数可以用同样的方式处理。

直方图以下的面积可以使用任何一对轴来求得:

- 水平的英寸尺度和纵向的每英寸的百分数尺度,或
- 水平的标准单位尺度和纵向的每标准单位的百分数尺度。

随便哪一种方式,得出的面积是百分数,因为纵轴采用了密度尺度。

上一章讲到对于许多数列,大致 68%的项在离平均数的 1 个

SD 以内。就是范围

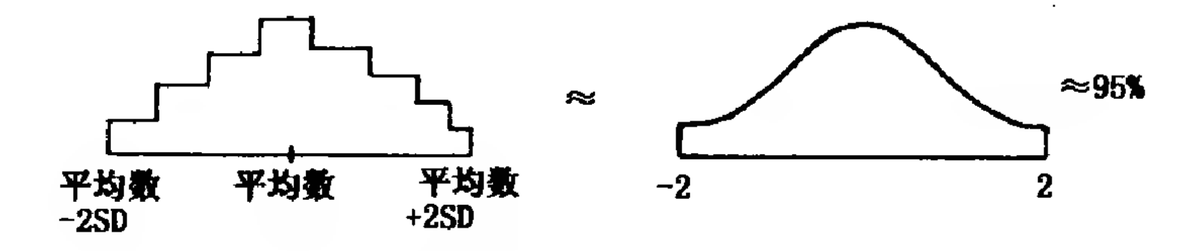
平均数−1 个 SD　至　平均数+1 个 SD。

为了知道 68%的出处,考察图 2。身高在平均数的 1 个 SD 范围内的妇女的百分数等于在平均数的 1 个 SD 以内的直方图下面的面积。该面积在图 2 中绘上阴影。直方图相当好地仿效正态曲线。它的某些部分比正态曲线略高一些,某些部分略低一些。然后高与低相抵。直方图下方的阴影面积恰好与曲线下方的面积差不多相等。在−1 与+1 之间的正态曲线下面的面积大约为 68%。那就是 68%的出处。

对于许多数列,大致 95%的项在离平均数的 2 个 SD 以内。就是范围

平均数−2SD　至　平均数+2SD。

理由是一样的。如果直方图遵循正态曲线,直方图下面的面积将大致与曲线下面的面积相等。在−2 与+2 之间曲线下面的面积大约为 95%。



于是,对于许多数列来说,在一个区间内的项的百分数可以估计如下。首先,将区间换算成标准单位;其次,求正态曲线下相应的面积。求取面积的方法将在第 2 节中讲述。最后,第 3 节将说明如何把这两步合在一起。整个过程称为正态近似。近似在于求面积之前用正态曲线代替原来的直方图。

习题 A

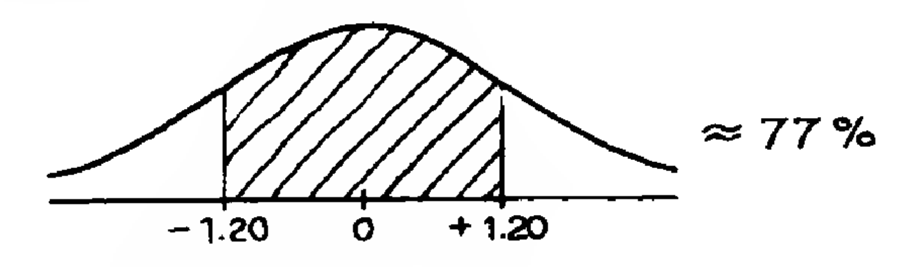
1. 某一次考试,各个得分的平均数是 50,SD 是 10。
   (a)将下列得分中的每一个换算成标准单位:60,45,75。
   (b)求得分,它们按标准单位是:0,+1.5,−2.8。
2. (a)将下列数列中的每一项换算成标准单位(那就是,利用数列的平均数和 SD):13,9,11,7,10。

(b)求换算了的数列的平均数和 SD。

这些习题的答案在第 679 页上。

2. 求正态曲线下的面积

在本书末，有一张表给出了正态曲线下的面积。例如。为求在 −1.20 与 1.20 之间正态曲线下的面积，在标 Z 的列中查阅 1.20 并且读出标面积的列中的项。这大约是 77%，因此在 −1.20 与 1.20之间正态曲线下的面积大约是 77%。



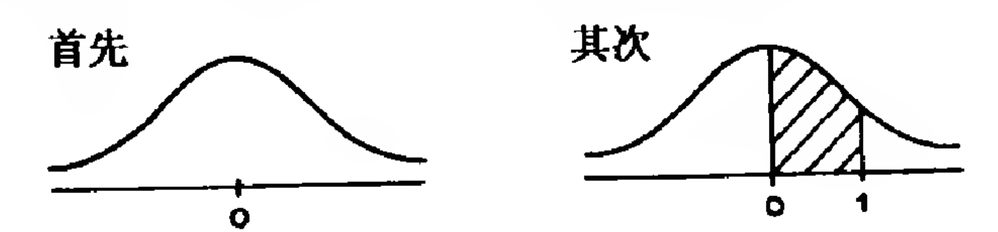
但是你也将需要求其他的面积。



求这样的面积的方法用例子来说明。

例 2. 求 0 与 1 之间正态曲线下的面积。

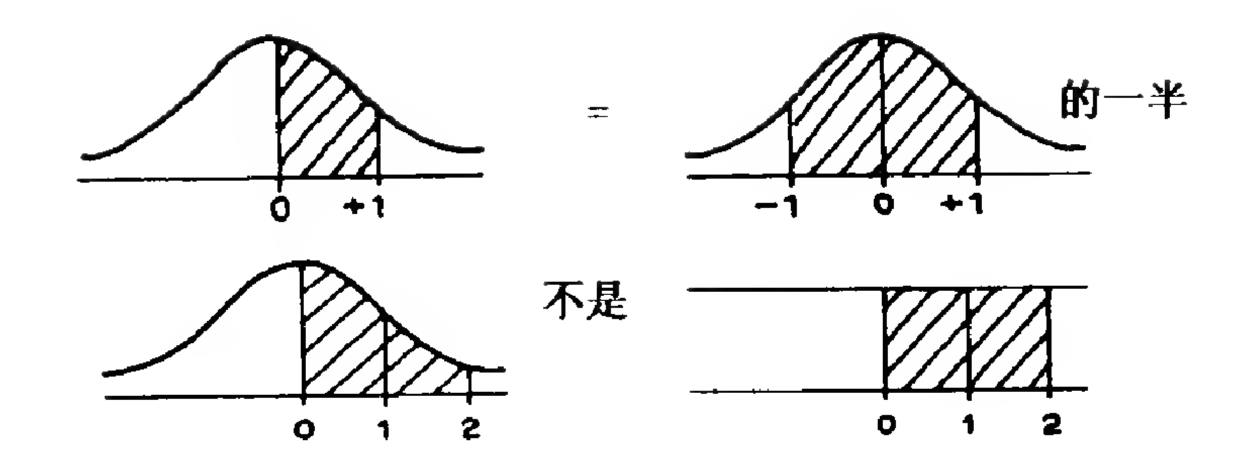
解：首先，作正态曲线的草图，然后在欲求的面积上绘以阴影。



表将给你 −1 与 +1 之间的面积。大约为 68%。由对称性，在 0 与 1 之间的面积是 −1 与 +1 之间面积的一半，那就是，$1/2\times 68\%=34\%$。

例 3. 求 0 与 2 之间正态曲线下的面积。

解：这不是在 0 与 1 之间面积的两倍，因为正态曲线不是矩形。

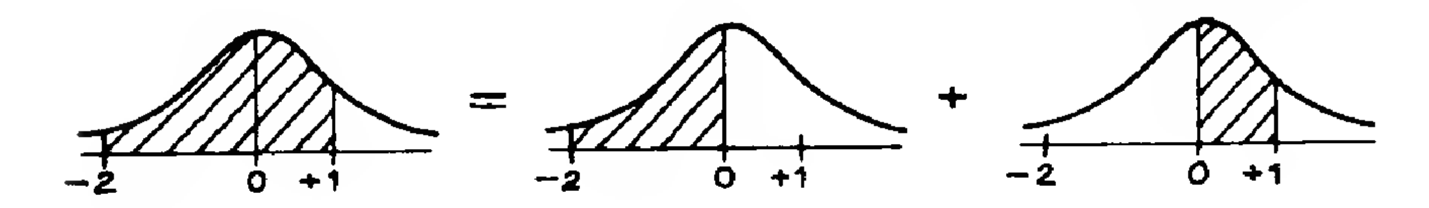


过程同例 2 中的一样。在 −2 和 2 之间的面积可以从表中找到。大约是 95%。由对称性，在 0 与 2 之间的面积是该数的一半：

$$\frac{1}{2}\times 95\%\approx 48\%。$$

例 4. 求 −2 与 +1 之间在正态曲线下的面积。

解　在 −2 与 1 之间的面积可以剖成两块其他形状的面积

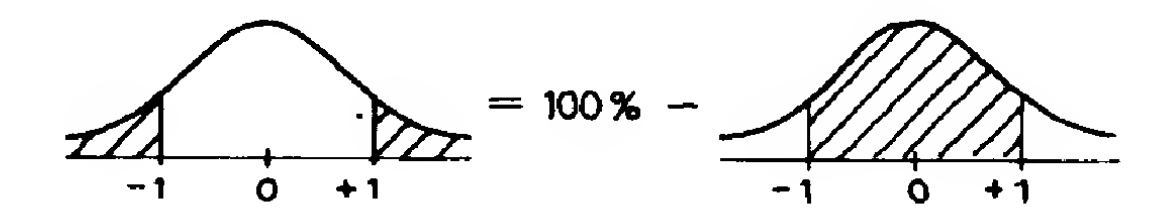


由对称性，在 −2 与 0 之间的面积等于在 0 与 2 之间的面积，大约为 48%。在 0 与 1 之间的面积大约是 34%。因此在 −2 与 1 之间的面积大约是

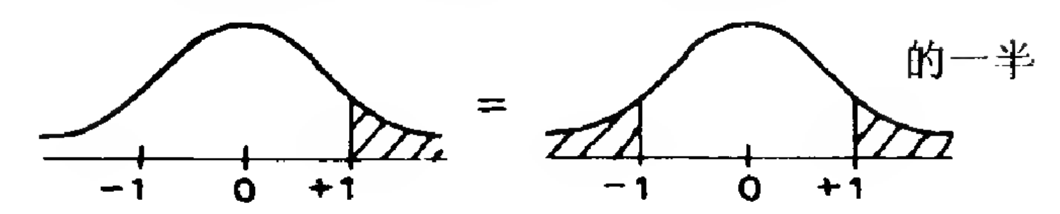
$$48\%+34\%=82\%$$

例 5. 求正态曲线下 1 右边的面积。

解：表给出在 −1 与 1 之间的面积，为 68%。因此在该区间外的面积是 32%。

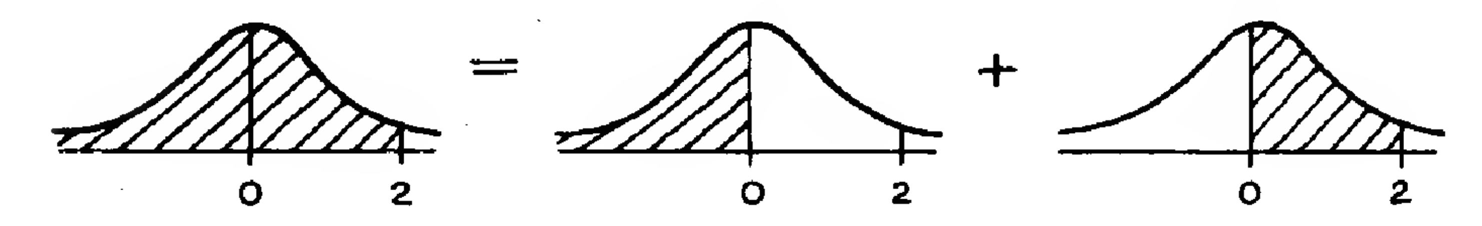


由对称性，1 右边的面积是该数的一半，或 16%。



例 6. 求正态曲线下 2 左边的面积。

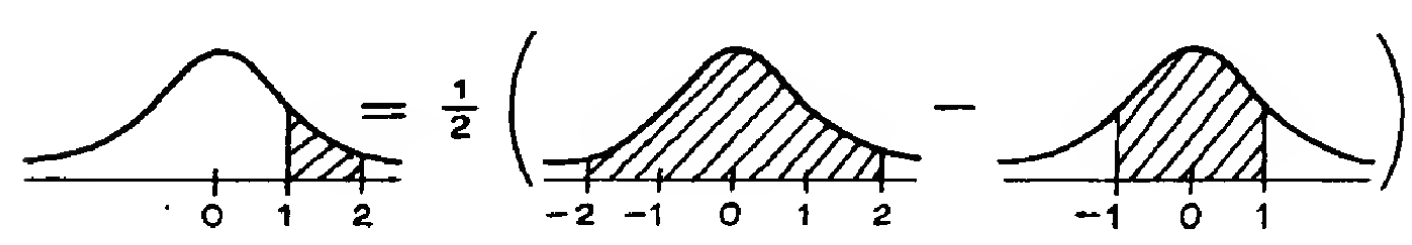
解　2 左边的面积是 0 左边的面积同 0 与 2 之间的面积之和。



由对称性，0 左边的面积是总面积的一半：$1/2\times100\%=50\%$。在 0 与 2 之间的面积大约是 48%。和为 $50\%+48\%=98\%$。

例 7. 求在 1 和 2 之间正态曲线下的面积。

解



在 −2 与 2 之间的面积大约是 95%；在 −1 与 1 之间的面积大约是 68%。其差的一半是

$$\frac{1}{2}\times(95\%-68\%)=\frac{1}{2}\times27\%\approx14\%。$$

没有一个固定的程序供解这类问题使用。问题是画一张把你想求的面积与表中可以查到的某些面积关联起来的图。

习题 B

1. 求正态曲线下的面积——

(a)1.25 的右边　　(b)−0.40 的左边

(c)0.80 的左边　　(d)−0.85 的右边

(e)－1.35 与 1.35 之间　　(f)0.40 与 1.30 之间

(g)－0.30 与 0.90 之间　　(h)－1.10 与－0.35 之间

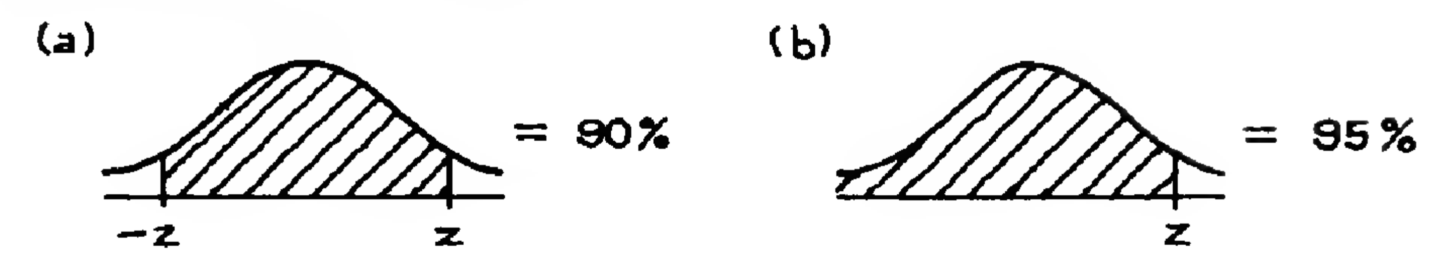
(i)－1.5 至 1.5 的外面

2. 填空：

(a)在±____之间正态曲线下的面积等于 68%。

(b)在±____之间正态曲线下的面积等于 75%。

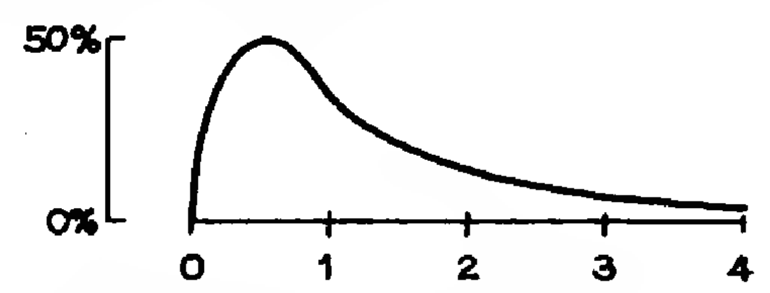
3. 正态曲线草绘在下面；求解 Z：



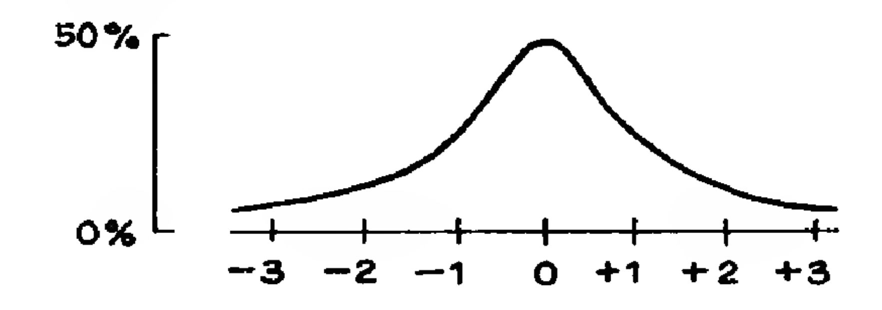
4. 某曲线(非正态)草绘在下面。在它下面的总面积是 100%，在 0 与 1 之间的面积是 39%。

(a)倘若可能，求 1 右边的面积。

(b)倘若可能，求 0 与 0.5 之间的面积。



5. 某曲线(非正态)草绘在下面。它关于 0 点对称，而且在它下面的总面积是 100%。在－1 与 1 之间的面积是 58%。



(a)倘若可能，求 0 与 1 之间的面积。

(b)倘若可能，求 1 右边的面积。

(c)倘若可能，求 2 右边的面积。

这些习题的答案在第 679—680 页上

3. 数据的正态近似

正态近似的方法在这里将用例子来讲解。图像看上去是如此简单以致于你可能认为它们不值得画出来。但是，这样容易失去想求的面积的线索。请画图像。

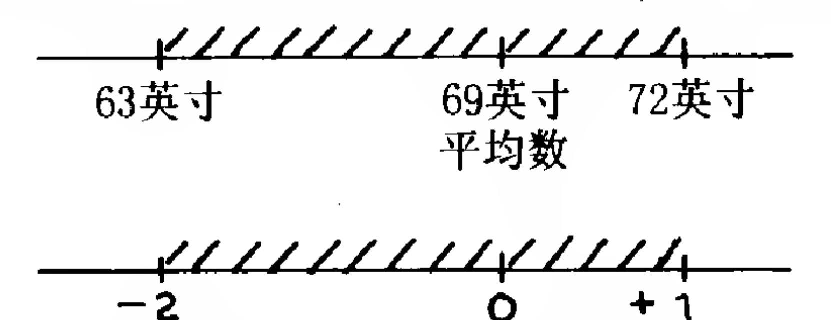
例 8. HANES 中 18—74 岁男子身高平均为 69 英寸；SD 是 3 英寸。使用正态曲线估计身高在 63 英寸与 72 英寸之间的这些男子的百分数。

解　百分数由（身高直方图下面）63 英寸和 72 英寸之间的面积给定。

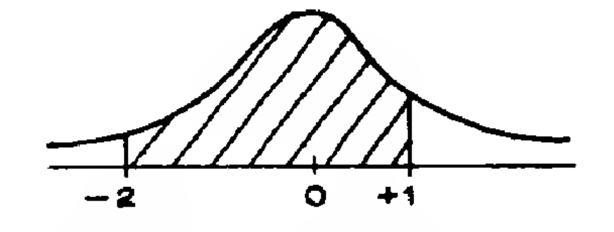
第一步，画一条数轴并对该区间涂以阴影。

63英寸　　　　72英寸

第二步，在轴上标出平均数。画一条新轴并换算成标准单位。

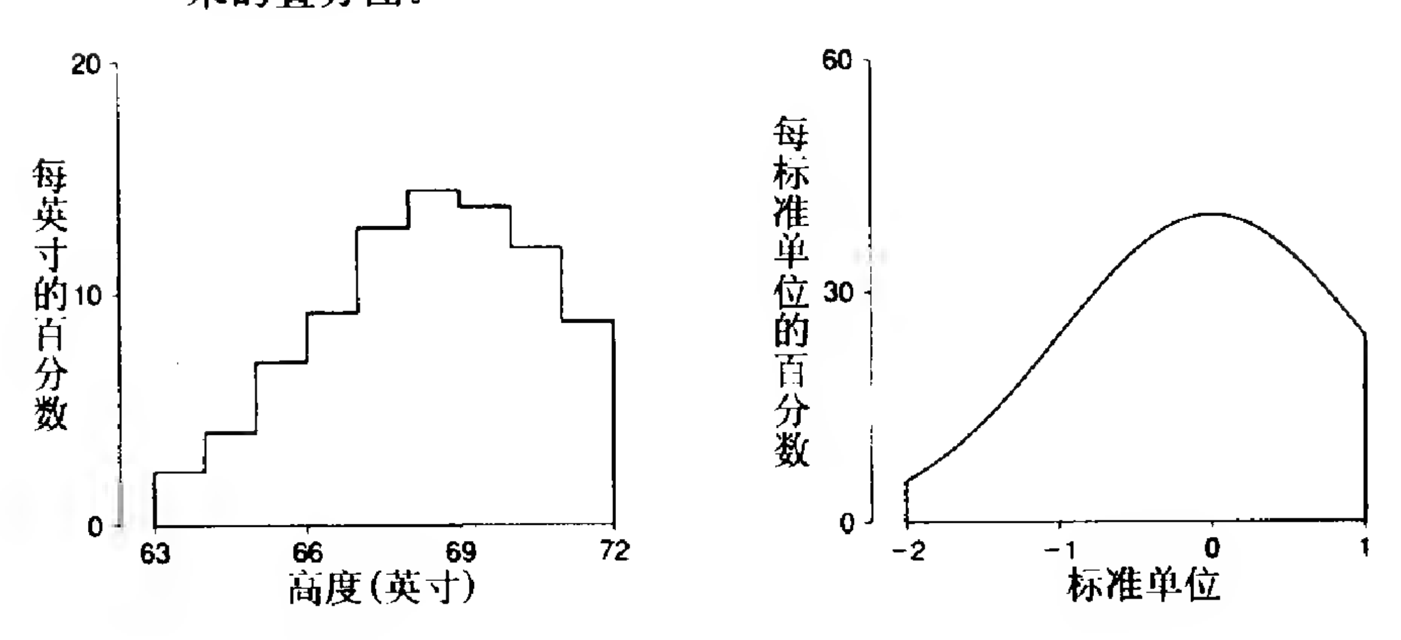


第三步，绘正态曲线简图，并求在第二步所得阴影的标准单位区间上方的面积。百分数近似地等于阴影部分面积，差不多是 82%。



使用正态曲线，估计大约有 82%的身高在 63 英寸与 72 英寸之间。这仅仅是一个近似，但是相当好：实际上 84%的男子处于那个范围。见图 3。

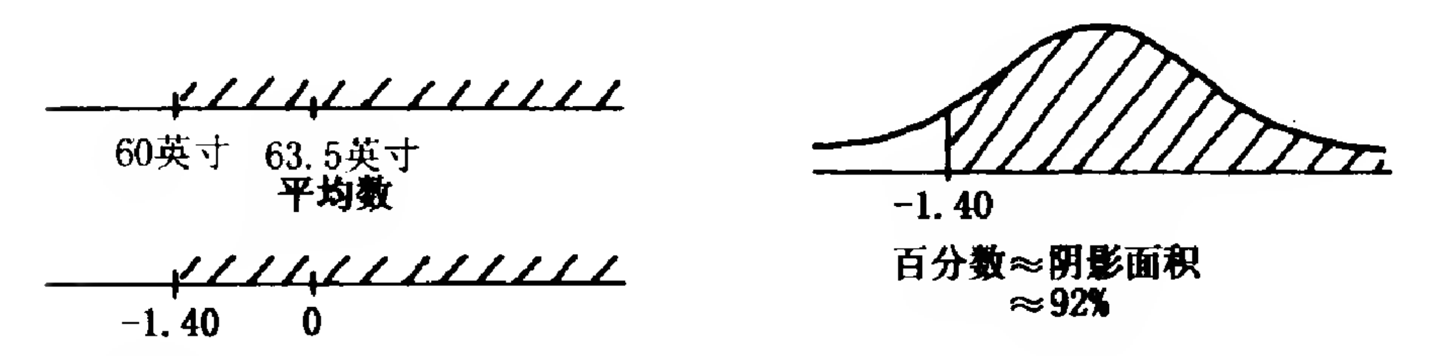
图 3 正态近似在于计算面积之前，用正态曲线代替原来的直方图。



例 9. HANES 中 18—74 岁妇女的身高平均为 63.5 英寸；SD 为 2.5 英寸。使用正态曲线估计身高在 60 英寸以上的百分数。

解　60 英寸的身高是平均数下 1.4SD：

$$(60-63.5)/2.5=-1.4$$

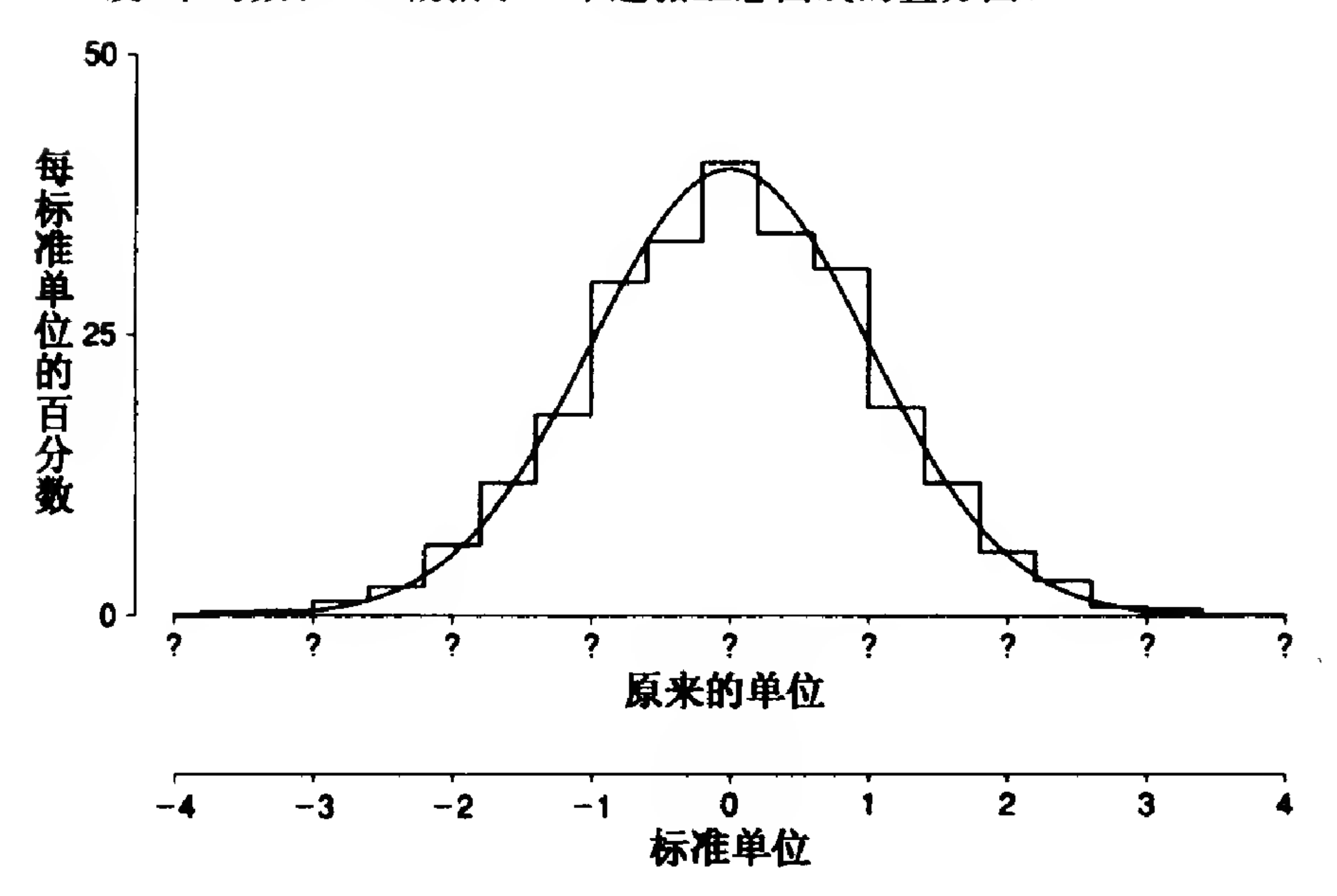


使用正态曲线，估计大约 92%的妇女身高超过 60 英寸。这个估计差不多是正确的。

一个值得注意的事实是许多直方图遵循正态曲线。(在第五部分继续描述。)对于这样的直方图，平均数与 SD 是好的概括统计量。如果一个直方图遵循正态曲线，它看来有点儿象图 4 中的草图。平均数决定了中心，而 SD 给出散布度。那几乎是关于直方图所要说的一切——假如它的形状象正态曲线的话。

然而，记住有许多其他直方图不遵循正态曲线的事实也是重要的，在这种情况，平均数和 SD 是不好的概括统计量。有关于此的更多内容在下一节中。

图 4 平均数与 SD。通过确定中心的位置和度量关于中心的散布度，平均数和 SD 概括了一个遵循正态曲线的直方图。



习题 C

1. 对于 HANES 中 18—24 岁的妇女，平均身高大约是 64.3 英寸，SD 大约是 2.6 英寸。使用正态曲线，估计妇女的百分数，她们的身高为——
   (a)66 英寸以下
   (b)在 60 英寸与 66 英寸之间
   (c)72 英寸以上
2. 对于这些妇女，平均体重大约是 134 磅；SD 大约是 27 磅。使用正态曲线，估计妇女的百分数，她们的体重为——
   (a)134 磅以下
   (b)在 100 磅与 150 磅之间
   (c)150 磅以上
3. 在法学院某班，进来的学生的 LSAT 分平均大约为 30；SD 为 8。LSAT 分的直方图很好地遵循于正态曲线。
   (a)班中大约多少百分数的学生得分低于 36?
   (b)某学生 LSAT 在平均数之上 0.5SD。大约多少百分数的学生具有比她所得的低一些的分数?

4. 图 2 中，身高在 61 英寸与 66 英寸之间的妇女的百分数精确地等于在＿＿＿＿下的面积，近似地等于在＿＿＿＿＿＿＿＿下的面积。

选择：

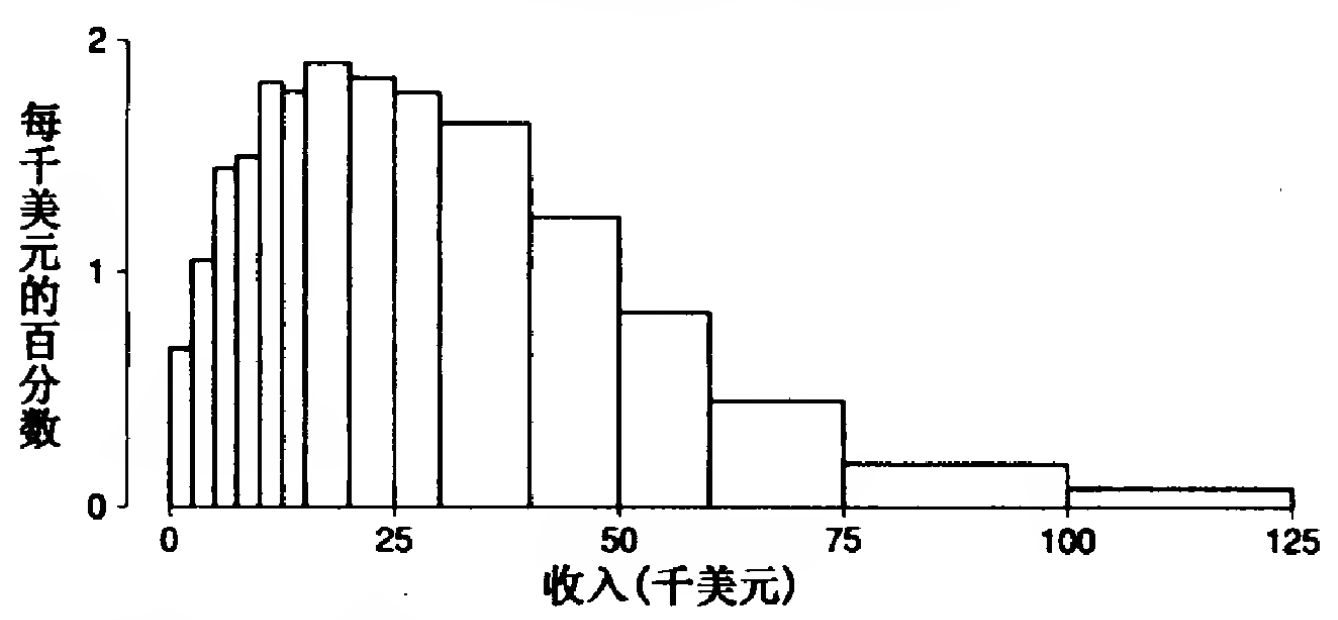
正态曲线　　　　　　直方图

这些习题的答案在第 680 页上。

4. 百分位数

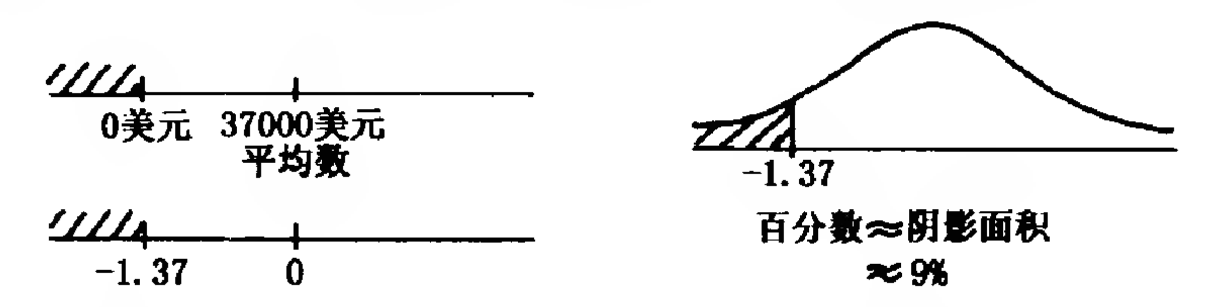
平均数和 SD 可以用来概括遵循正态曲线的数据。对于其它类型的数据，它们是不令人满意的。取 1987 年在美国的家庭收入分布，如图 5 所示。

图 5　1987 年美国按收入的家庭分布。



来源：现场人口调查数据磁带。

图 5 中家庭的平均收入大约是 37 000 美元；SD 大约是 27 000 美元。因此正态近似使人想到大约 9％的家庭具有负的收入：



铸成大错的理由：图 5 中的直方图根本就不遵循正态曲线；它有一个长的右尾部。为了概括这样的直方图，统计学家常常使用百分位数(表 1)。

表 1 1987 年在美国关于家庭收入挑选的百分位数

1	1300 美元
10	8800 美元
25	17100 美元
50	30800 美元
75	48000 美元
90	68500 美元
99	125600 美元

注：按照人口调查规定，有关和无关的子家庭排除在该表之外。

来源：现场人口调查数据磁带。

收入分布的第 1 百分位数是 1 300 美元，意思是指大约 1%的家庭具有 1 300 美元或更少一些的收入，而 99%的家庭具有该水平以上的收入。第 10 百分位数是 8 800 美元：大约 10%的家庭具有低于那个水平的收入，而 90%的家庭在该水平之上。第 50 百分位数恰好是中位数（第四章）。

由定义，四分位数间距等于

第 75 百分位数—第 25 百分位数。

当 SD 对位于分布尾部的小百分数事例给予过多的关注时，四分位数间距有时候用作散布的度量。对于表 1，四分位数间距是 30 900 美元。

出于自身的原因，统计学家称 de Moivre 曲线是“正态的”，这产生一个印象即其它曲线是变态的。事实不是这样。许多直方图遵循正态曲线相当地好，而许多其它的——象收入直方图——不是如此。在本书的后面部分，我们将提出一类数学理论，它将有助于解释什么时候直方图会遵循正态曲线。

习题 D

1. 表 1 中收入在 10 000 美元与 70 000 美元之间的家庭的百分数大约是_________。

60%　　75%　　95%

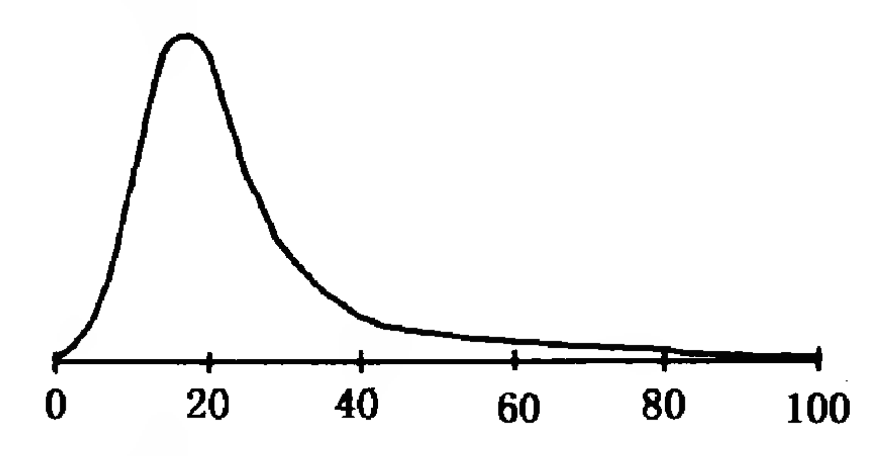
2. 使用第 39 页上表 1 中的数据，1973 年家庭收入分布的第 25 百分位数大约是 7 000 美元，10 000 美元，还是 25 000 美元？
3. 皮肤褶厚度用于测量体内脂肪。一个有关皮肤褶厚度的直方图给在下面；

水平轴上的单位是毫米(mm)。皮肤褶厚度的第 25 百分位数是________ 25mm。使用下面的措词之一填入空格。或者这能从图象来确定吗?

相当地小于

大约

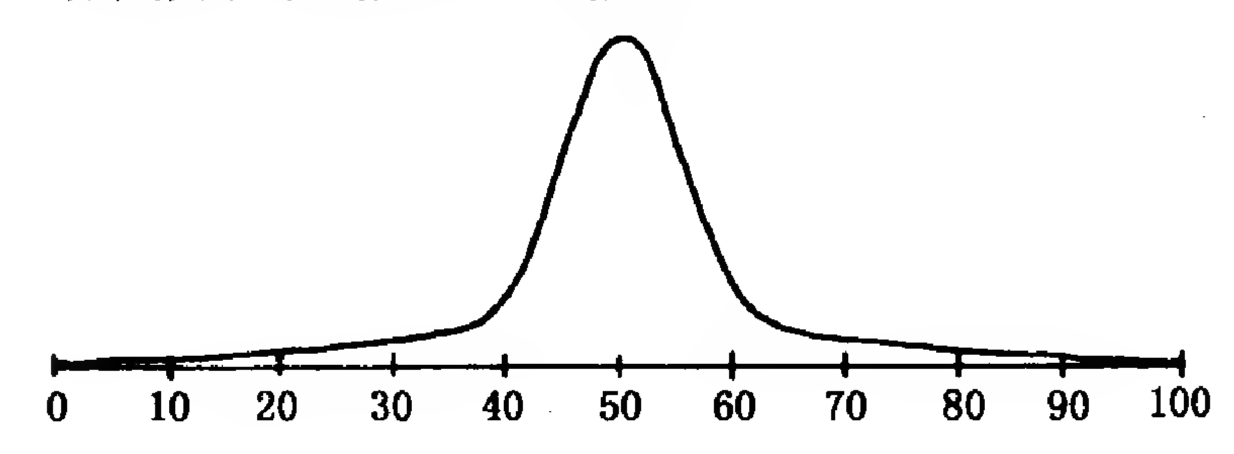
相当地大于



4. 一个直方图草绘在下面。

(a)它与正态曲线怎样不同?

(b)四分位数间距大约是 15,25,还是 50?



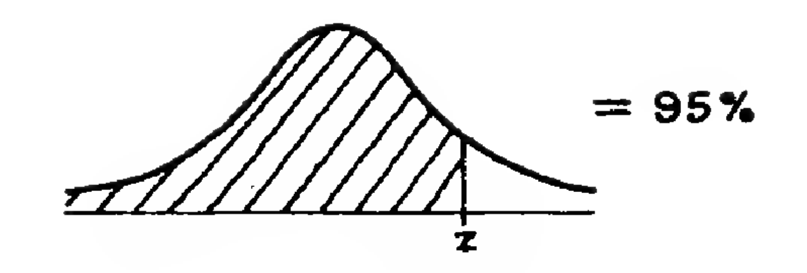
这些习题的答案在第 680 页上。

5. 百分位数和正态曲线

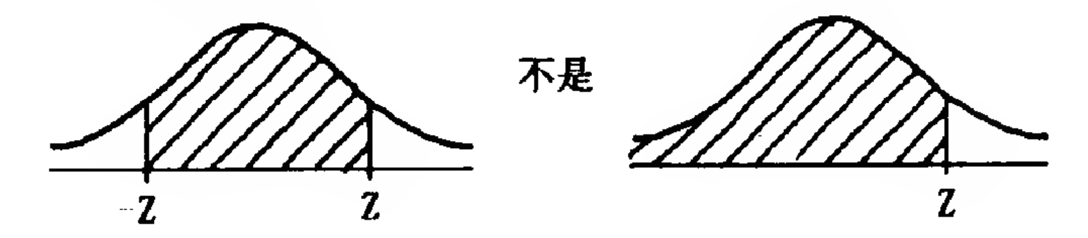
当直方图遵循正态曲线时,正态表可以用来估计它的百分位数。用例子来讲解方法。

例 10. 某一年在某大学的所有申请者中,数学 SAT 得分平均为 535,SD 是 100,并且得分遵循正态曲线。估计得分分布的第 95 百分位数。

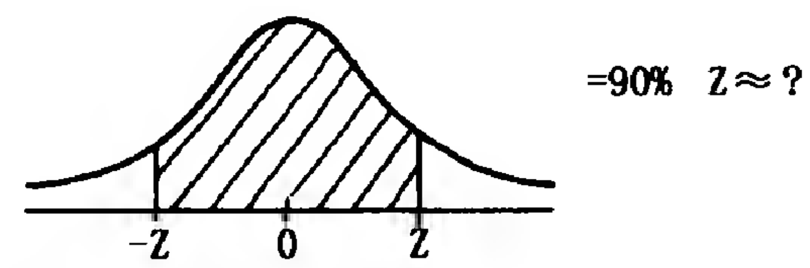
解:这个得分位于平均数之上一定量的 SD。我们需要求那个量;称它为 Z。有一个关于 Z 的方程:



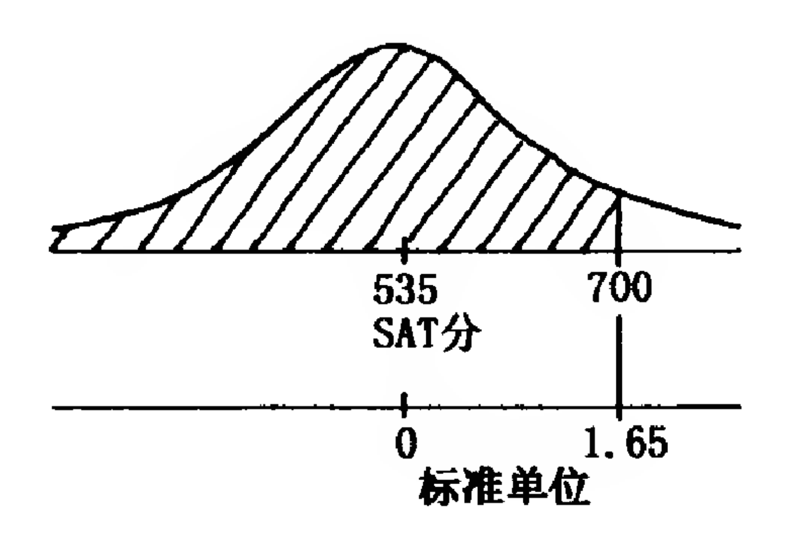
正态表不能直接使用，因为它给出在$-Z$与Z之间的面积，而不是Z左边的面积：



我们的Z右边的面积是5%；因此在$-Z$左边的面积也是5%。那末在$-Z$与Z之间的面积必定是$100\%-5\%-5\%=90\%$。



由表，$Z\approx1.65$。你必须将平均数之上1.65SD认为是数学SAT得分分布的第95百分位数。换回到分数，那是在平均数之上大约$1.65\times100=165$分。得分分布的第95百分位数是$535+165=700$。



习题 E

1. 在例10中的大学里，一名申请者在数学SAT中得750分。（在所有申请者中间）她的百分位数序是什么？

2. 对于例 10 中的大学，估计数学 SAT 得分的第 90 百分位数。

3. 对于伯克莱新生，平均 GPA（等级分平均数）大约是 3.0；SD 大约是 0.5。直方图遵循正态曲线，估计 GPA 分布的第 30 百分位数。

这些习题的答案在第 680 页上。

6. 复习题

复习题可能包含前面几章的内容。

1. 下面的考试分数单具有平均数 50 和 SD10：

   39 41 47 58 65 37 37 49 56 59 62 36 48

   52 64 29 44 47 49 52 53 54 72 50 50

   (a)利用正态近似估计在平均数 1.25SD 范围以内的得分个数。

   (b)有多少个得分真正地在平均数 1.25SD 范围内？

2. 你正在察看计算机打印出来的 100 个考试成绩，它们已经被换算成标准单位。粗略地看一下头 10 项，你看到

   −6.2　3.5　1.2　−0.13　4.3　−5.1　−7.2　−11.3　1.8　6.3

   有什么错误吗？

3. 一名研究人员有一计算机文件，显示了某研究中 1 000 名对象的家庭收入。这些数据跨程从 1 年 5 800 美元至 1 年 98 600 美元。由于偶然事故，文件中最高收入变成为 986 000 美元。

   (a)这影响平均数吗？如果影响的话，相差多少？

   (b)这影响中位数吗？如果影响的话，相差多少？

4. 从六十年代中期至八十年代中期，在学业能力考试(SAT)成绩中存在着缓慢且稳定的下降。对于数学 SAT，1967 年的平均数大约是 492；到了 1987 年，平均数下降到大约 476。SD 几乎为常数 100，得分的直方图遵循正态曲线。下降 16 分也许看来不算多，但是对于分布的尾部有大的影响。

   (a)估计 1967 年得分超过 700 分的学生的百分数。

(b)估计 1987 年得分超过 700 分的学生的百分数。

(SAT 分数跨程从 200 至 800。似乎不是 SAT 变得较难;六十年代的大多数下降被认为来自于参加考试的学生总体变化的结果;七十年代的下降不能那样解释[②]。)

5. 在数学 SAT 方面,男子有明显的优势[②]。例如,在 1988 年男子在数学 SAT 方面平均大约 500 分,而妇女平均大约 455 分。两者的直方图均遵循正态曲线,且两个 SD 均大约是 100。

(a)估计 1988 年在该项考试中超过 600 分的男子的百分数。

(b)估计 1988 年在该项考试中超过 600 分的妇女的百分数。

6. 从参加 1988 年数学 SAT 考试的男子中随机地选出一名,你必须猜测他的考试成绩。如果你在不超过 50 分范围内猜中,你将获得 1 美元。

(a)你应当猜多少?

(b)你赢的机会是多少?

7. 在 1988 年一所法学院的申请者中间,平均 LSAT 得分大约是 40,SD 大约是 8,最高得分是 48。LSAT 得分遵循正态曲线吗? (LSAT 得分跨程从 10 至 48;全部所有考试参加者的平均数大约是 30,SD 大约是 7。)

8. 对于公共卫生总署研究中现抽烟的男子,每天所吸烟的平均支数大约是 27;SD 大约是 20 支。(直方图在第 47 页的上半部分给出。)如果你用正态近似去估计每天吸 2 支烟或不到 2 支烟的百分数,估计将会差不多正确呢,还是相当地过高,或者相当地过低?

9. 按照人口调查下的定义,一个"家庭"由 2 个或 2 个以上生活在一起的有亲戚关系的人所组成;一"户"有 1 个或 1 个以上的人居住在同一住房单元。1987 年,户平均收入比家庭平均收入少大约 10%。这可能是怎么回事? 简短地讨论之。

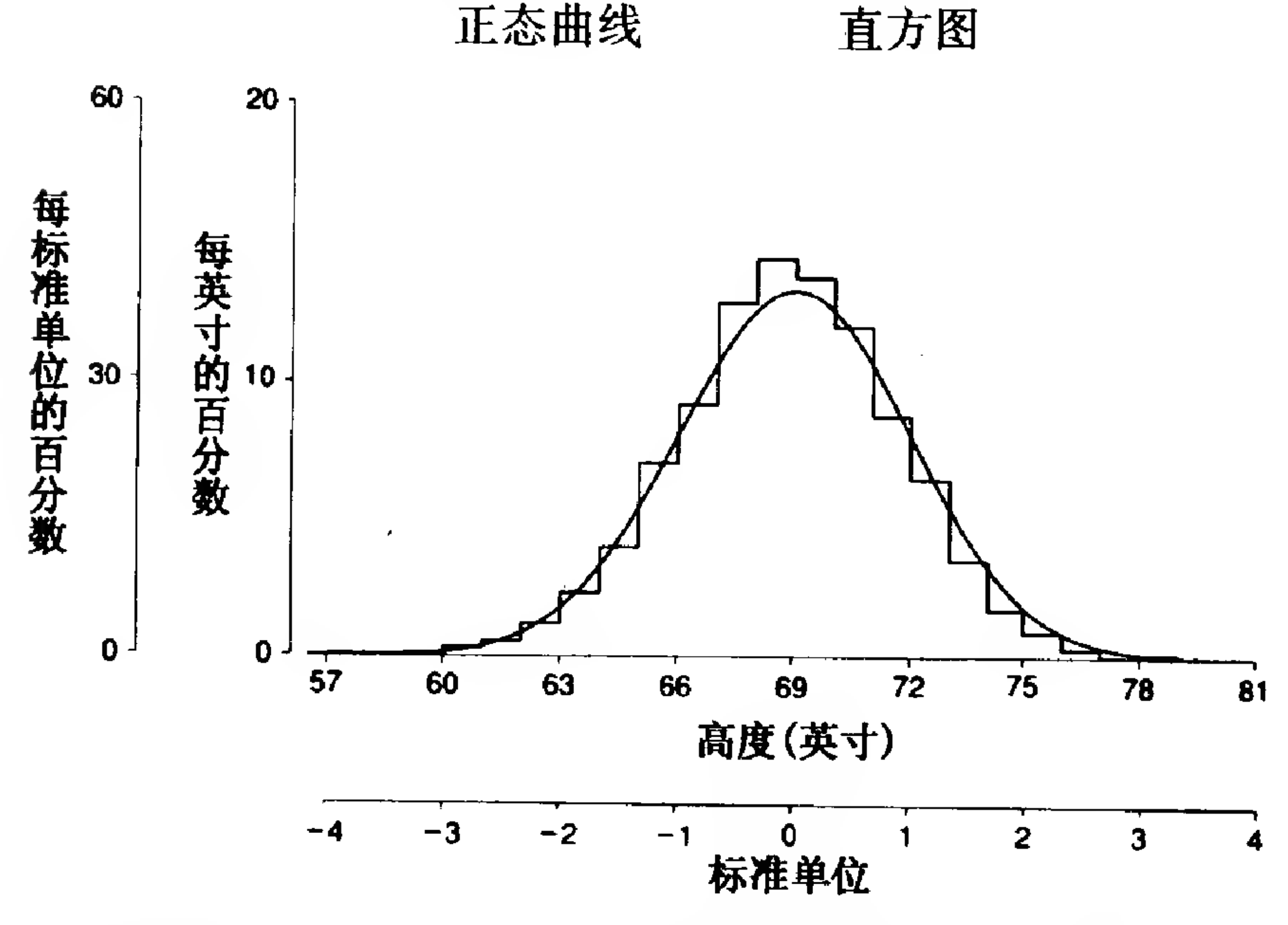
10. 在 HANES 中,18—74 岁的男子有 69 英寸的平均身高和 3 英寸的 SD。直方图在下面,并附有一条正态曲线。身高在 66 英

寸与72英寸之间的男子的百分数精确地等于在___(a)___与___(b)___之间___(c)___下的面积。这个百分数近似地等于在___(d)___与___(e)___之间___(f)___下的面积。对于(a),(b),(d),和(e),你的选择是

66英寸　　72英寸　　-1　　1

对于(c)和(f),你的选择是

正态曲线　　　　直方图



11. 下面的哪一个是正确的? 错误的? 解释或者给出例子。

(a)任何数列的中位数与平均数总是靠得很近。

(b)数列中的一半总是在平均数以下。

(c)对于一组有代表性的大的样本,直方图必定相当接近地遵循正态曲线。

(d)如果两个数列有完全相同的平均数50和相同的SD10,那末对这两个数列来说,在40与60之间的项的百分数必定完全相同。

12. 某学期,California大学Berkeley分校修统计学2的大约700名学生被询问他们修了多少门不同于统计学2的大学数学课

程。课程的平均数大约是 1.1;SD 大约是 1.5。关于数据的直方图看来是象(i),(ii),还是(iii)?



7. 小结

1. 正态曲线关于 0 点对称,在它下面的总面积是 100%。

2. 标准单位说出了一个值在平均数之上(+)或下(−)多少个SD。

3. 许多直方图具有与正态曲线大致相同的形状。

4. 如果一列数遵循正态曲线,落在一个给定的区间内的项的百分数可以通过将区间换算成标准单位,然后求正态曲线下相应的面积而估计。这个过程称为正态近似。

5. 一个遵循正态曲线的直方图可以根据它的平均数和 SD 相当好地重新构建;在这种情况,平均数和 SD 是好的概括统计量。

6. 所有直方图,不管它们是否遵循正态曲线,可以使用百分位数来概括。



"瞧FRED!这似乎是同一件事,是经过概括的。"

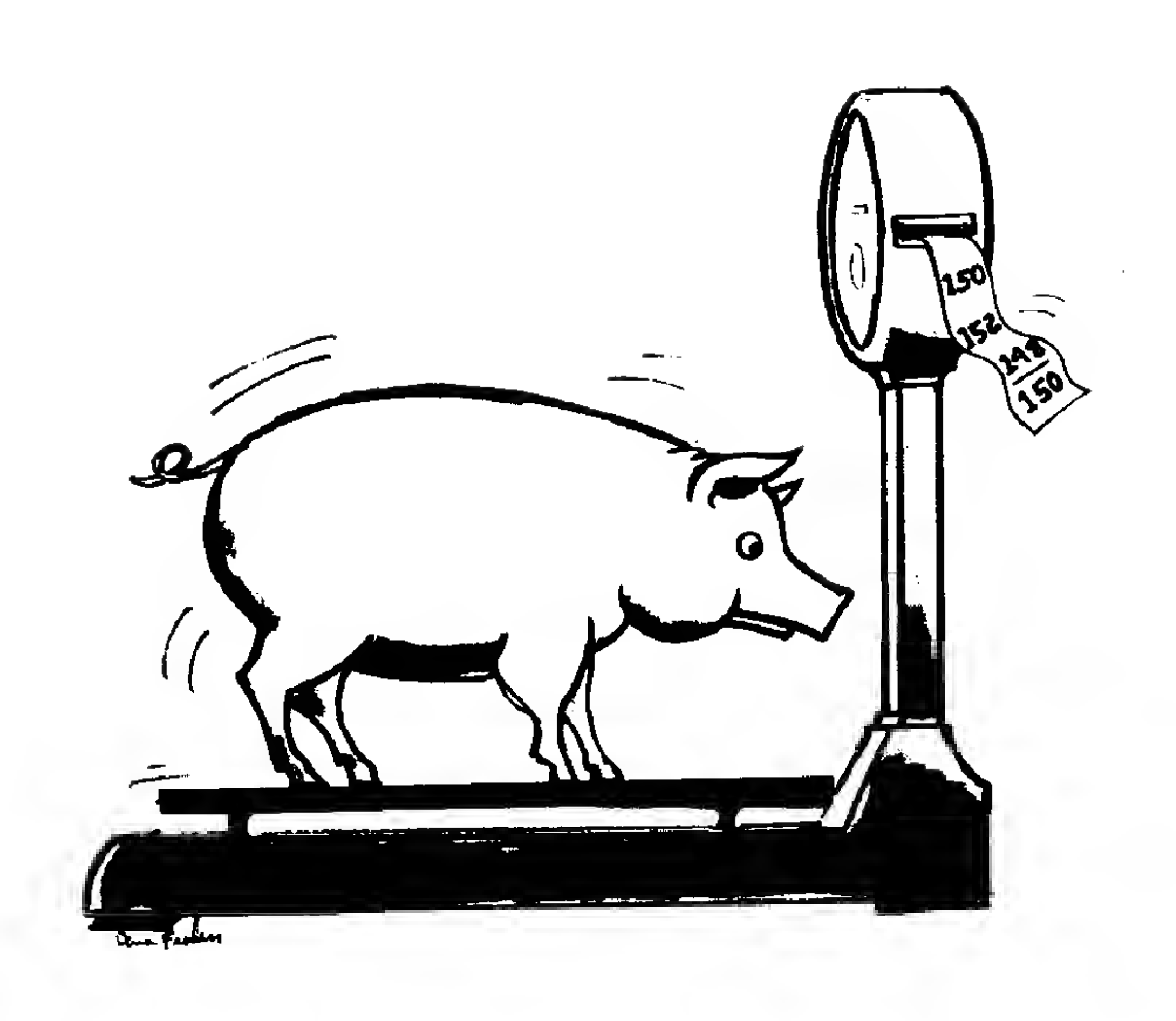
6
测量误差

耶稣：我来证明真理。
Pilate：真理是什么？

1. 引言

如果同一样东西测量若干次，在理想的世界里每次将得到相同的结果。但在实践中是存在差异的。每一次结果都产生机会误差，误差随一次次的测量而变化。处理这个问题最早的科学家之一是丹麦天文学家 Tycho Brahé(1546—1601)。但是可能在市场上当商人称香料和量丝绸的长度时首先注意到这个问题。即使在今天，贸易者不得不与机会误差相处。例如在一些畜牧场，猪是按重量出售的。猪站在磅秤上称若干次，每次的重量和它们的平均数打印在纸带上。价格是按照平均数来定的，因为采用平均使得机会误差至少在某种程度上抵消了。

存在着若干有关机会误差的问题：它们从哪里来的？可能有多大？在平均数中可能抵消掉多少？第一个问题是容易的：在大多数场合没有人知道。第二个问题稍后在本章中处理，而第三个问题将在第七部分中得到回答。



2. 机会误差

本节将讨论在华盛顿国家标准局做的精确性称量中的机会误差①。首先，简略地说明一下标准砝码。商店在磅秤上称商品。磅秤由县度量衡官员使用县标准砝码定期地检验。县标准也必须定期地校准(以外部的标准为对照来检验)。这在州水平上完成。而州标准由在华盛顿的美国国家标准局以国家标准为对照加以校准。

这一连串的比较终止于国际标准千克(为简短起见，千克)，这是一个保存在巴黎附近的国际计量局里的铂一铱合金砝码。按照国际公约(米突公约，1875)，“1 千克”定义为这个砝码在标准条件②下的重量。所有其它称量单位相对于千克而确定。例如，某物件假如其重恰好略少于 1 千克的一半则称为重 1 磅。更精确一点。

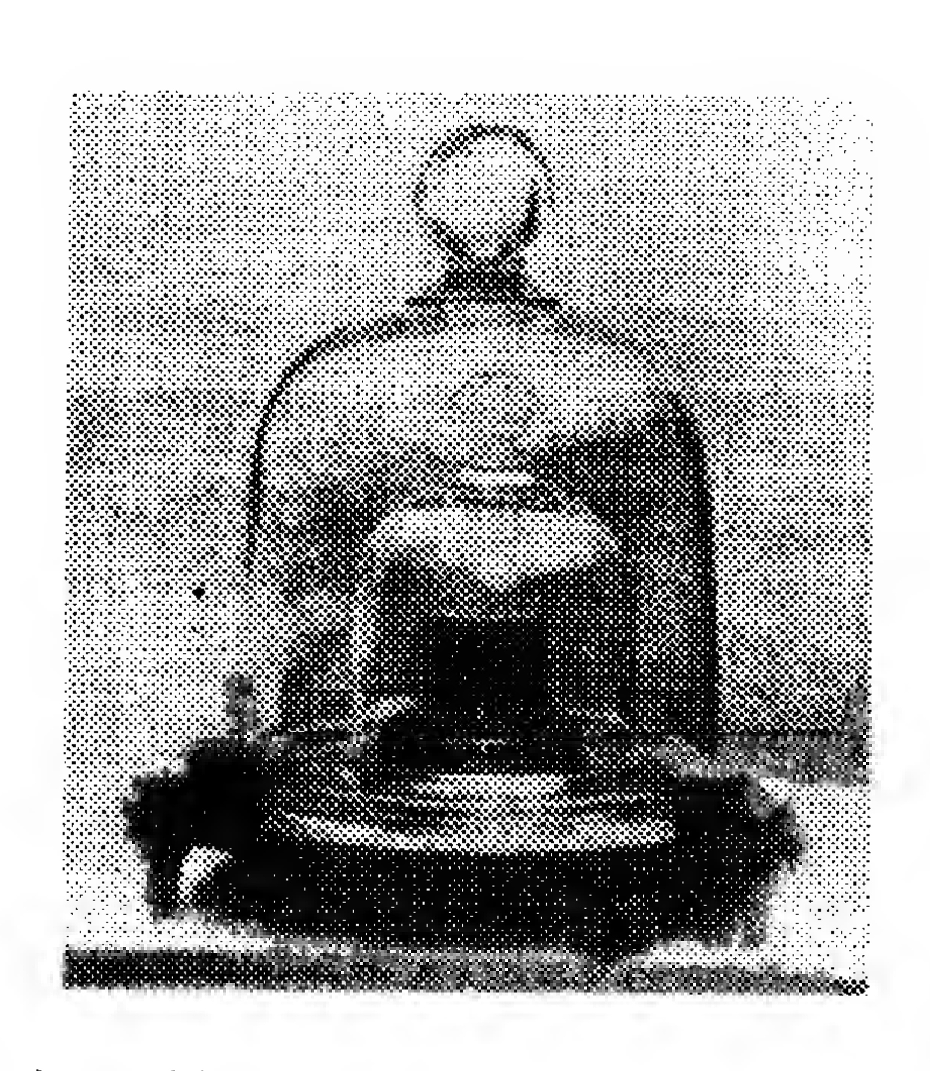
1 磅＝0.4539237 千克。

说一袋黄油重 1 磅，意即经过某些漫长而复杂的系列比较，它已经同在巴黎的千克相联系且为千克的 0.4539237 倍重——加减

某机会误差。

每一个签署米突公约的国家得到一个国家标准千克，它的精确重量相对于千克尽可能精确地确定。这些原型按抽签的方式分发，美国得到千克＃20(K_{20})。所有的美国国家标准值相对于K_{20}而确定。

图1 美国国家标准千克，K_{20}



来源：国家标准局公告(1905)。

在美国，超级市场的称重的精确性最终依赖于在标准局中所做的校准工作的精确性。一个基本问题是复制性：如果重复测量，它将改变多少？标准局通过对他们自己的一些砝码的重量作重复的确定而了解这个问题。我们将讨论一个这样的砝码的结果，它称为NB10因为它为国家标准局(National Bureau)所拥有并且其名义上的值是10克——两个五美分镍币的重量。(一袋黄油有"名义上的"重量1磅；精确的重量将有一点儿不同——包装中的机会误差。同样地，制造NB10的人企图使它重10克，差那么一点点而未

能达到。)

NB10 由标准局在 1940 年左右获得,从那时起直至现在他们大约每星期称一次。我们将观察 1962—1963 年在标准局由 Almer 先生和 Jones 夫人完成的 100 次称量。这些测量是在同一房间相同的设备上进行的。尽一切努力使得每次测量遵循相同的程序。所有已知的影响结果的因素,象气压或者温度,尽可能保持为常数。

在该系列数中头 5 次称量是

9.999591 克

9.999600 克

9.999594 克

9.999601 克

9.999598 克

第一眼的感觉,这些数看来都相同。可是观察更仔细些,只是头 4 个数字固定在 9.999。后 3 个数字是不稳定的;它们随着不同次的测量而改变。这就是机会误差在起作用[③]。

NB10 比 10 克略轻一些。因此取代每次写出 9.999,标准局仅仅报告 NB10 不足 10 克的量。对于第一次称量,即为

0.000409 克

那么多 0 会使人分散注意力,因此标准局不以克工作而以微克:1 **微克**是 1 克的百万分之一。用这些单位,关于 NB10 的头 5 次测量值容易读出。它们是

409　400　406　399　402

所有 100 次测量值在表 1 中给出。往下看表,你能看到这些结果在 400 微克左右徘徊,有多有少。最小的是 375 微克(#94);最大的是 437 微克(#86)。在其间有很多变化。为使脑海中有所印象,一微克是一大点灰尘的重量;400 微克是一或二粒细盐的重量。这真正地是精确称量!

即使如此,不同的测量不可能全都正确。NB10 不足 10 克的精确量未必等于表中的第一个数,或者第二个数,或者它们中的任

意一个数。事实上，尽管作了这100次测量的努力，NB10的精确重量仍然不知道，也许是不可能知道的。NB10是在10克以下400微克左右——多少有点小误差。

表1 关于NB10的100次测量值，由Almer和Jones在国家标准局进行。单位是10克以下的微克。

编号	结果	编号	结果	编号	结果	编号	结果
1	409	26	397	51	404	76	404
2	400	27	407	52	406	77	401
3	406	28	401	53	407	78	404
4	399	29	399	54	405	79	408
5	402	30	401	55	411	80	406
6	406	31	403	56	410	81	408
7	401	32	400	57	410	82	406
8	403	33	410	58	410	83	401
9	401	34	401	59	401	84	412
10	403	35	407	60	402	85	393
11	398	36	423	61	404	86	437
12	403	37	406	62	405	87	418
13	407	38	406	63	392	88	415
14	402	39	402	64	407	89	404
15	401	40	405	65	406	90	401
16	399	41	405	66	404	91	401
17	400	42	409	67	403	92	407
18	401	43	399	68	408	93	412
19	405	44	402	69	404	94	375
20	402	45	407	70	407	95	409
21	408	46	406	71	412	96	406
22	399	47	413	72	406	97	398
23	399	48	409	73	409	98	406
24	402	49	404	74	400	99	403
25	399	50	402	75	408	100	404

可能仍然有点儿奇怪的是标准局一遍又一遍地重复称同一个砝码。该做法的目的之一是质量控制：如果关于NB10的测量值从10克以下400微克跳到10克以上500微克，那么某些事情出了差错而且需要加以修整。（为此理由，NB10称为检验砝码，可以用

来检验称量过程。)

为了看到重复测量的另一个用途,设想一个科学实验室将一个名义上 10 克的砝码送到标准局去校准。当然,由于机会误差的缘故,一次测量不能最终定局。实验室希望知道机会误差可能有多大。

有一个直接的途径去查明:将同一个砝码送回去作第二次称量。如果两次结果相差少量微克,那末在每一次中的机会误差大小仅仅可能是少量微克。另一方面,如果两次结果相差大约几百微克,那末每次测量都有可能相差几百微克。对于 NB10 的重复称量省却了不止一次地递送砝码的麻烦。不必去要求重复校准,因为标准局已经做了这项工作并且发现测量值有多好。

> 不管多么小心地去做,得到的测量值可能有点不同于真值。如果重复一次,将显现一点不同。相差多少? 在信赖任何一次测量值之前,必定要面对这个问题。回答的最佳办法是通过重复测量。

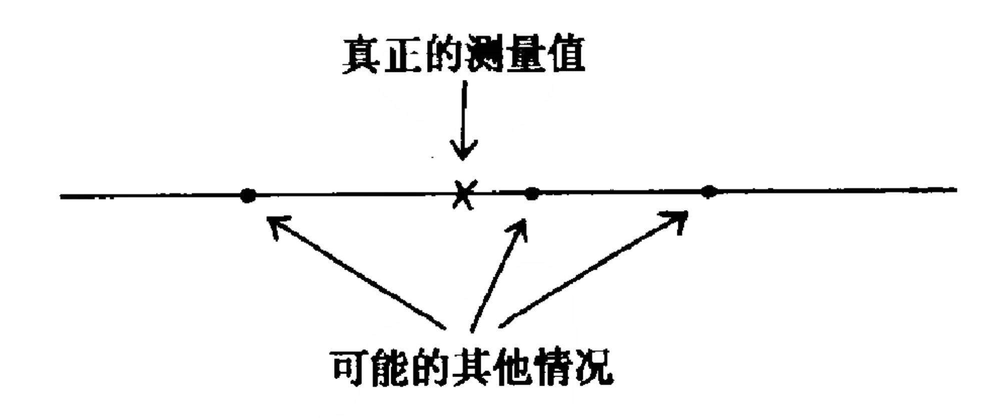


表 1 中 100 次测量值的 SD 仅是 6 微克多一点。SD 告诉你关于 NB10 的每一次测量产生的机会误差大约为 6 微克。机会误差在大小上为 2 或 5 或 10 微克左右是相当普遍的。机会误差大约是 50 或 100 微克必定相当罕见。结论:用相同的过程校准其他的 10 克砝码。机会误差应当是 6 微克左右。

一系列重复测量的 SD 是单次测量中机会误差可能大小的估计。

有一个等式有助于说明原因：

单独测量值＝精确值＋机会误差。

机会误差使每一次单独的测量值偏离精确值。偏离大小随着不同次的测量而变化。重复测量中的变化性反映了机会误差的变化性；两者均以 SD 来测定。

更细致地着手这个问题，表 1 中的所有 100 个测量值的平均数是 10 克以下 405 微克。很可能接近于 NB10 的精确重量。但是表 1 中第一个测量值离平均数相差 4 微克：

409－405＝4。

因此第一个测量必定离精确重量几乎 4 微克。那时，机会误差几乎是 4 微克。类似地，第二个测量值在平均数以下 5 微克，它的机会误差必定是－5 微克左右。距平均数的代表性偏差大小约为 6 微克，因为 SD 是 6 微克。因而，典型的机会误差必定象 6 微克那样的大小。

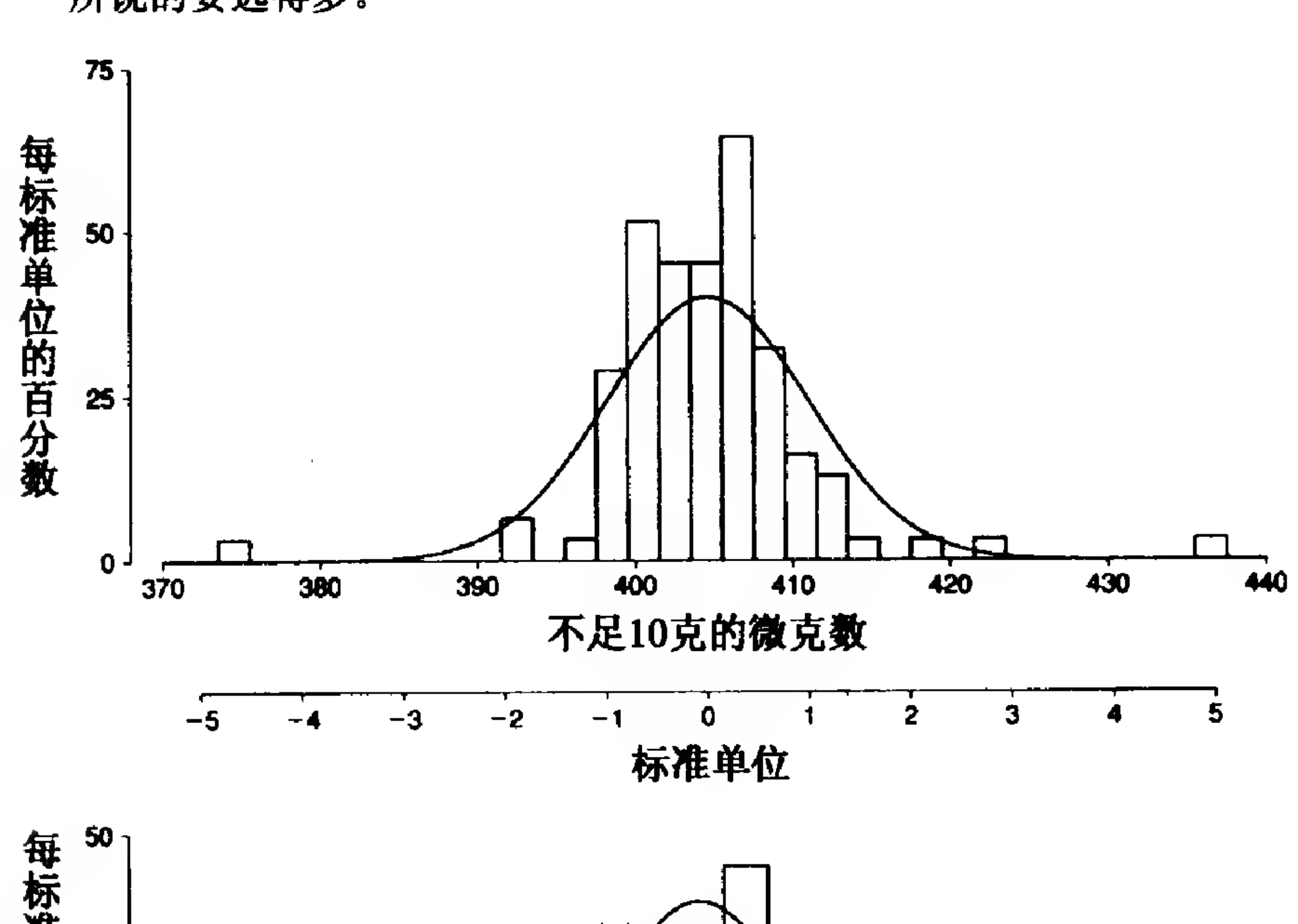
3. 离群点

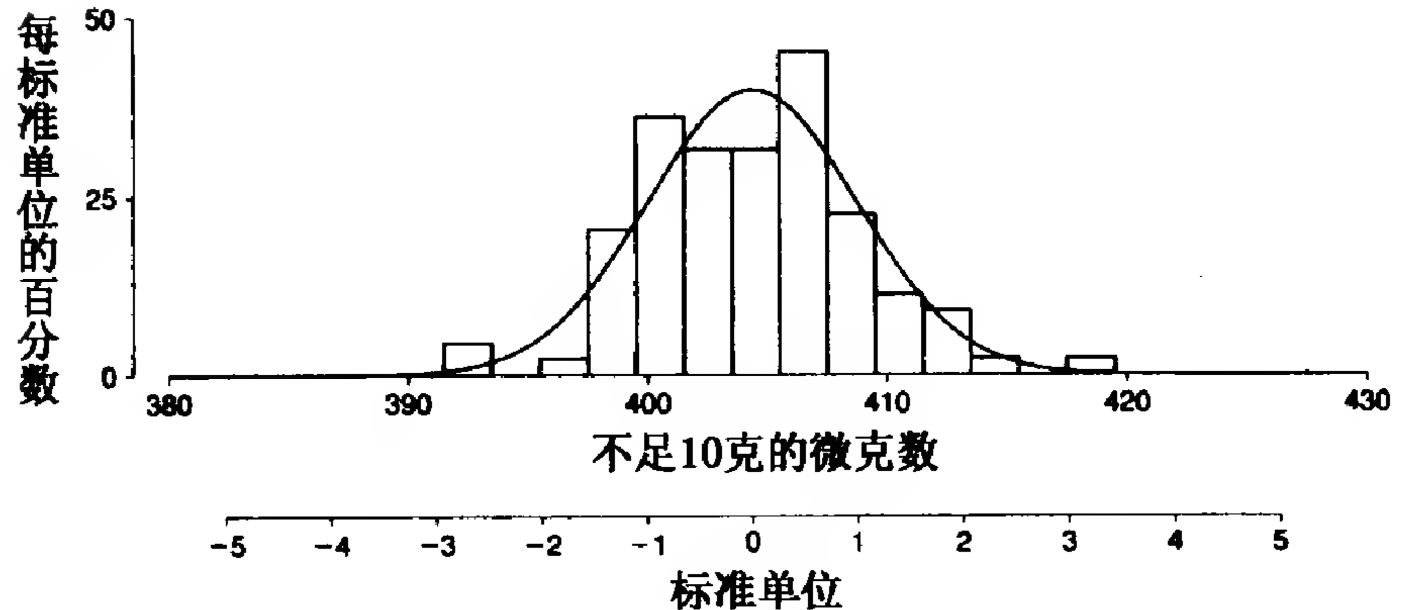
表 1 中所载的测量值拟合正态曲线有多好？回答是，不十分好。测量值 # 36 离平均数 3SD；# 86 和 # 94 为 5SD 远——较少的非凡事例。这样极端的测量值称为离群点。它们不是大错而造成的。就标准局所告知，在得到这 3 个观察值时没出什么差错。但是，这 3 个异常数据对 SD 有实质性贡献。因此，落入离平均数 1SD 以内的百分数为 86％——比起由正态曲线所预测的 68％来相当地多。见图 2(下一页)。

摒弃这 3 个离群点，余下的 97 个测量值平均为 10 克以下 404 微克，具有仅为 4 微克的 SD。因此平均数改变不多而 SD 下降了大约 30％。如图 2 所示，余下的 97 个测量值完全合理地遵循正

态曲线。总之，数据中的大多数遵循正态曲线，具有大约 4 微克的 SD。而少量的测量离开平均数比起正态曲线所说的要远得多。整体情况下的 6 微克 SD 是在直方图主要部分的 SD——4 微克——与离群点之间的折衷。

图 2 离群点。上面部分给出了关于 NB10 的全部 100 次测量值的直方图；为比较起见画了一条正态曲线。曲线拟合得不好。第二部分表示了去除 3 个离群点后的数据。曲线拟合得较好。数据中的大多数遵循正态曲线，可是少量的测量值离平均数比正态曲线所说的要远得多。





在细微的测量工作中，预期有较少百分数的离群点。NB10 数据的唯一不寻常之处是报告了离群点。这里给出标准局对于不报告离群点而不得不说的话[④]。由于是官方口气，语调相当严厉。

> 在统计方法对于测量数据分析的应用中，一个主要困难是获得合适的数据。问题是比较经常地有意识或者也许无意识地试图按照人们所乐意的那样构造一个特殊的实施过程，而不是接受真正的事实…。基于任意的对实施过程的限制而拒绝数据会严重曲解对实际过程变化的估计。这种做法使得原先计划的目的失效。实际性能参数需要接受全部数据，它们不能由于某种原因而被摒弃。

当研究者看到一个离群点时需要作出一个艰难的选择。要么省略它；要么不得不认可他们的测量值不遵循正态曲线。曲线的声誉如此之高使得第一种选择成为通常的一种选择——理论战胜了经验。

4. 偏性

假如一个屠夫将他的姆指放在磅秤上一起秤一大块肉。那会引起测量中的误差。可是几乎没有什么偶然性。举另外一个例子。假如布店使用一根布卷尺来测量，该布卷尺已从 36 英寸拉伸到 37 英寸长。他们卖给顾客的每一"码"布增加了额外的 1 英寸。这不是机会误差——因为它总是偏向顾客。屠夫的姆指和拉伸的卷尺是偏性或者系统误差的两个例子。

偏性以相同的方式影响所有测量值，将它们推向同一方向。机会误差随着不同次的测量而变化，有时候向上有时候向下。

如果在测量过程中没有偏性，大量重复测量值的平均数将给出待测物体的精确值；所有的机会误差将抵消。但是，当偏性出现时，大量测量的平均数将要么过高要么过低。由于偏性，就象机会

误差那样，每一个测量值会偏离真值：

单独测量值＝精确值＋偏度＋机会误差

通常，仅仅观察测量值本身是没有什么办法察觉偏性的。要么它们必须与客观标准比较要么与理论预测比较。在美国，所有称重依赖于 K_{20} 与千克之间的联系。这两个砝码已经许多次地被比较，估计 K_{20} 比千克极微小地轻一点——相差十亿分之十九。作为一种补偿，在标准局的所有重量计算均向上调正大约十亿分之十九。然而，这个因子可能恰好有少量的偏离——毕竟，它也是一些测量过程的结果。在美国所有的称重系统地相差相同（微小的）百分数。这是偏性的另一个例子——但是无人为此烦恼。

5. 复习题

这些习题覆盖了第一部分和第二部分的所有内容。

1. 正确还是错误，并解释："一个正在使用可得到的最好设备的有经验的科学家只需要测量物体一次——假如他不出差错。最终，假如他测量同一物体两次，他将在两次得到相同的结果。"
2. 一个木匠正在使用卷尺量取木板的长度。
   (a)偏性的一些可能来源是什么？
   (b)哪一个更易于出现偏性，是钢卷尺还是布卷尺？
   (c)布卷尺的偏度将随时间而变吗？
3. 正确还是错误，并解释——
   (a)偏性是一种机会误差。
   (b)机会误差是一种偏性。
   (c)测量值通常受到偏性与机会误差两方面的影响。
4. 在健康检查的调查（类似 HANES，但在 1960—1961 年进行）中，有 6 672 名对象。每名对象的性别在调查的两个不同阶段被记录下来。在 17 例中，存在不一致之处：对象在一次访问中被记录为男性，而在另一次访问中则被记录为女性。你如何说明原因？

5. 你将一支码尺送到当地实验室去校准，要求重复验校 3 次。他们报告了下述数值：

35.96 英寸　　36.01 英寸　　36.03 英寸

如果你将码尺送回去进行第四次校准，你将预期得到 36 英寸 ±________

0.01 英寸左右　　0.03 英寸左右　　0.06 英寸左右

6. 在一个初等统计学教程中的 19 名学生被要求使用一根读数可到千分之一英寸的游标尺去测量一个桌面的厚度。每人测量两次，结果如下所示（单位为英寸）。例如，第一个人两次测量得到 1.317 和 1.320 英寸。

(a)这些学生互相独立工作吗？

(b)你的一些朋友不相信机会误差。你如何能使用这些数据去说服他们？

	测量值（英寸）			测量值（英寸）	
人	第 1 次	第 2 次	人	第 1 次	第 2 次
1	1.317	1.320	11	1.333	1.334
2	13.26	13.25	12	1.315	1.317
3	1.316	1.335	13	1.316	1.318
4	1.316	1.328	14	1.321	1.319
5	1.318	1.324	15	1.337	1.343
6	1.329	1.326	16	1.349	1.336
7	1.332	1.334	17	1.320	1.336
8	1.342	1.328	18	1.342	1.340
9	1.337	1.342	19	1.317	1.318
10	13.26	13.25			

7. 利用下面的选择填空，并给出例子以证明你选择了正确的答案。

(a)一数列的 SD 是 0。这意思是说____。

(b)一数列的 r,m,s 是 0。这意思是说____。

选择:

(i)数列中没有数。

(ii)数列中所有的数相同

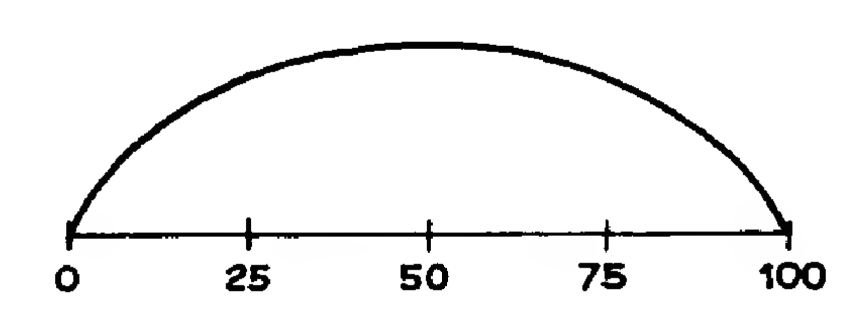
(iii)数列中所有的数为 0。

(iv)数列的平均数是 0

8. 在某课程中,期终考的得分遵循正态曲线;平均数大约是 50,SD 大约是 10。大约多少百分数的学生得分在 35 与 70 之间?

9. 在上题从班中随机地选择一个学生,你必须猜测他的得分。如果你的猜测正确到 10 分之内你将赢 1 美元。你应当猜什么并且赢的可能有多大?

10. 在某课程中,期终得分的直方图看起来象下面的草图。正确还是错误:因为这不象正态曲线,必定在考试中存在某些差错。请解释原因。



11. 在某大班中,男生平均身高为 5 英呎 9 英寸;他们的 SD 是 3 英寸。女生平均身高为 5 英呎 3 英寸;她们的 SD 是 3 英寸。在该班中有 150 名男生和 150 名女生。将男女生放在一起:

(a)平均身高将为____。

(b)身高的 SD 将略大于 3 英寸,恰好为 3 英寸,还是略小于 3 英寸?

解释你的推理。

12. 重复习题 11,此时在该班中有 200 名男生和 100 名女生。(男女生各自的平均数与 SD 仍不变。)

13. 为了表明赞成增加审判员的工资，Warren E. Berger 作证说“比起 1789 年宪法开始直至 1969 年这段时期”，更多的法官在任期中离开了首席法官的职位。(纽约时报，1988 年 12 月 9 日。)连首席法官都犯了统计错误。Berger 论据的错误是什么？

14. 为测定运动对于心脏病危险的效果，研究人员决定对伦敦两个大组公交人员——驾驶员与售票员——比较发病率。售票员在终日走动卖票时得到大量的锻炼。这两个组的年龄分布非常相象，而且所有的对象从事于同一工作有至少 10 年的一段时期。在售票员中间的心脏病发病率事实上略低一些[5]。这在多大程度上说明运动减少了心脏病的危险？

15. 在一些司法案件中，有审判前的准备会议，在会上法官与双方律师协商调解案件或者至少在审判之前确定问题的性质。观察资料提出了审判前准备会议促进了解决以及加快了审判过程，但是存在着疑虑。

在新泽西法院，审判前的准备会议是强制性的。然而，在七个县做了一次实验：在六个月的期间内，2 954 件人身伤害案例(主要是汽车事故)被随机地指派进处理组或对照组。对于 1 495 件对照案例(A 组)，审判前准备会议仍是强制性的。对于 1 459 件处理案例，准备会议是可任意选择的——任一方律师可以提出请求。在处理案例中间，701 例选择了审判前的准备会议(C 组)。758 例没有选择(B 组)。

分析资料的研究人员留神去发现是否审前准备会议促进了案例在审前解决；或者，如果他们进行审判，准备会议是否缩短了审判时间。(这很要紧，因为审判时间非常昂贵。)

研究人员报告主要结果如下。(表中材料引自他的报告[6]。)

(i)审判前的准备会议对审前解决没有影响；B 组与(A+C)组具有相同的百分数。

	B 组	(A+C)组
进行了审判	22%	23%
案例数	701	2079

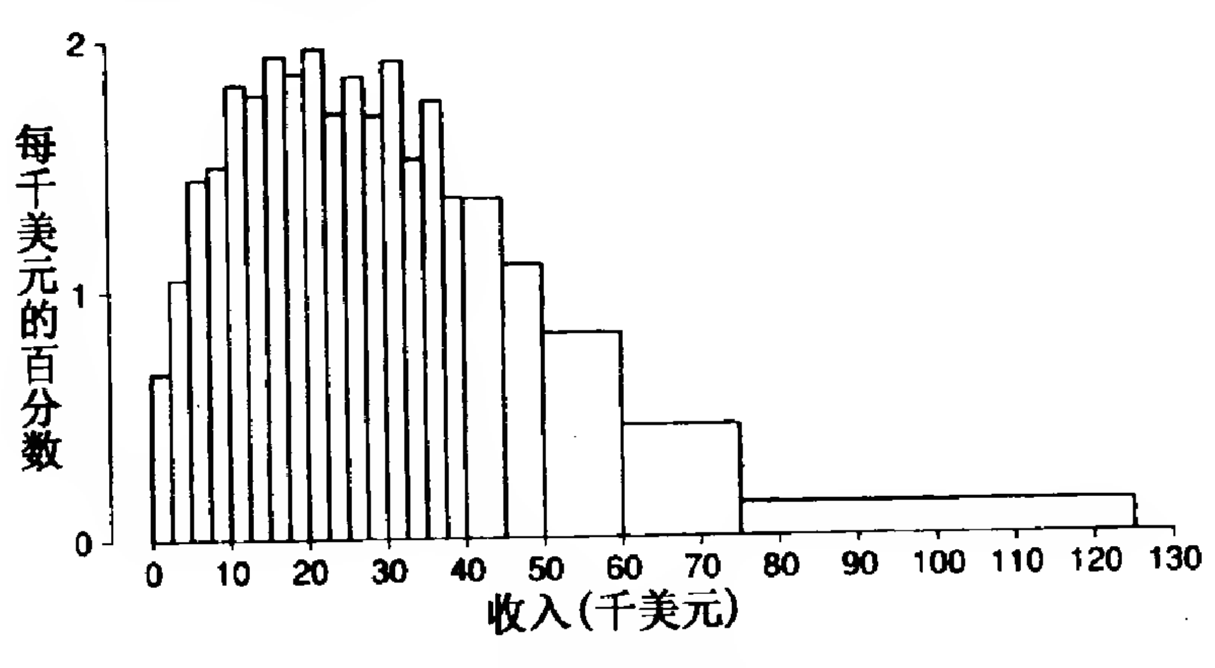
(ii)审判前的准备会议不缩短审判时间；在拒绝审判前准备会议的案例中短期审判的百分数是最高的。

在进行审判的案例中间审判时间的分布

	B 组	A 组	C 组
审判时间(以小时计)			
1. 5 或更少	43%	34%	28%
2. 5 以上至 10	35%	41%	39%
3. 10 以上	22%	26%	33%
案例数	63	176	70

简短地评价该分析。

16. 三月份，现场人口调查要求一大组有代表性的美国人说出他们在过去的一年内的收入[7]。1987 年有关家庭收入的直方图如下所示。(小组区间包含左端点但不包含右端点。)在 10 000 美元与40 000美元之间，块形有规则地从高到低相互交叉。那是为什么？请简短地解释。



17. 在 1976—1980 年期间，公共卫生总署在健康和营养调查研究中研究了一大组有代表性的美国人[8]。左撇子回答者的百分数随年龄而平稳递减，从 20 岁的 10%减少到 70 岁 4%。“数据表

明随着年龄的老化许多人从左撇子改变为使用右手。”正确还是错误？为什么？

6. 小结

1. 不管多么小心地去做，得到的测量值可能与以前的有点儿不同。这反映了机会误差。在研究人员信赖一个测量值之前，他们应该估计误差的可能大小。最好的做法是通过重复测量。

2. 单次测量中机会误差的可能大小可以通过在同等条件下的一系列重复测量值的 SD 来估计。

3. 偏性，或者系统误差，引起某过程所得到的测量值有系统地过高或者过低，公式是

单独测量值＝精确值＋偏度＋机会误差。

机会误差随不同次的测量而改变；偏性则不是这样。

4. 即使在仔细的测量工作中，少量百分数的离群点可能出现。

5. 平均数和 SD 可能受到异常数据的强烈影响。因而直方图将根本不遵循正态曲线。

7
点和线的描绘

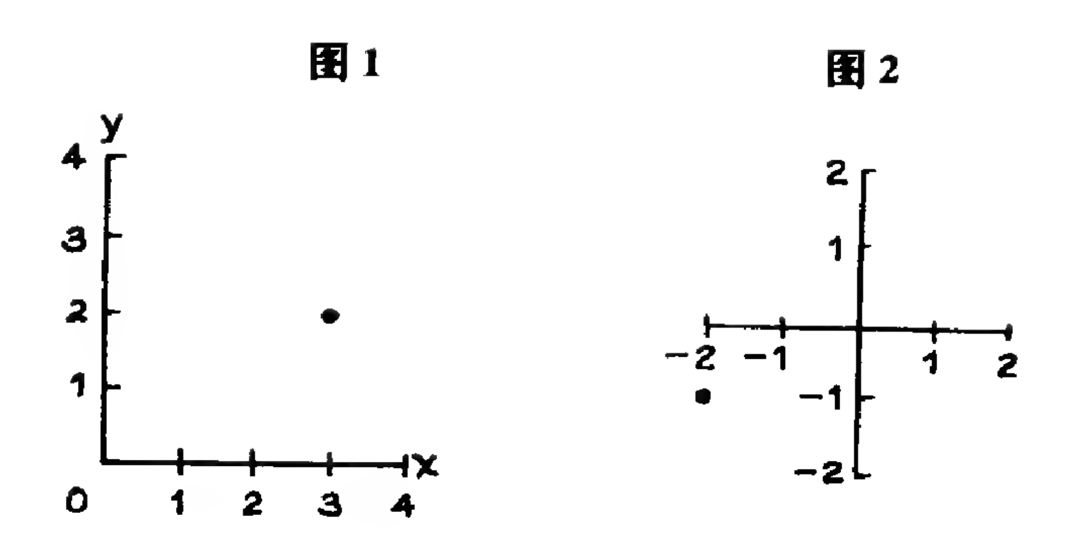
问：点对线意味着什么？
答：线归结为点。

1. 从图中辨认点

本章复习一些描绘点和线的有关概念，本书第三部分将用到这些概念。读者可现在就阅读本章，亦可在第三部分遇到困难时再回头读它。如果现在就阅读本章，前四节最为重要，最后一节较难。

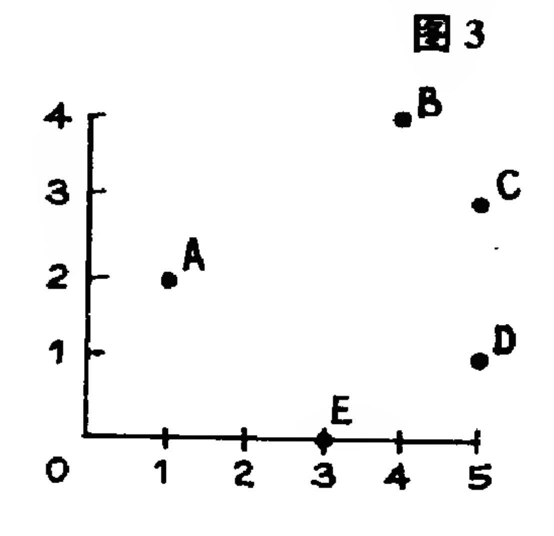
图 1 给出了一条横轴（x 轴）和一条纵轴（y 轴）。图中点的横坐标为 3，因为它与 x 轴上的 3 成一纵直线；其纵坐标为 2，因为它与 y 轴上的 2 成一水平直线。这个点可书写为 $x=3, y=2$。有时可进一步缩写为(3,2)。图 2 中的点为(−2,−1)，因为它在 x 轴上 −2 的正下方，在 y 轴上 −1 的正左方。

用成对的数来代表点的思想是 René Descartes（笛卡尔，法国 1596—1650）提出来的。为了纪念他，x 和 y 坐标常称为“笛卡尔坐标”。



习题 A

1. 图 3 有 5 个点，写出每个点的 x 和 y 坐标。
2. 在图 3 中，如果从点 A 移到点 B，x 坐标将增加____；y 坐标增加____。
3. 图 3 中某点的 y 坐标比点 E 的 y 坐标大 1，这是哪一个点？



这些习题在的答案第 680—681 页上。

2. 描点

图 4 给出一直角坐标系。为了描出点(2，1)，先找到 x 轴上的 2，该点将在其正上方，如图 5；然后找到 y 轴上的 1，该点将在其正右方，如图 6。

图 4

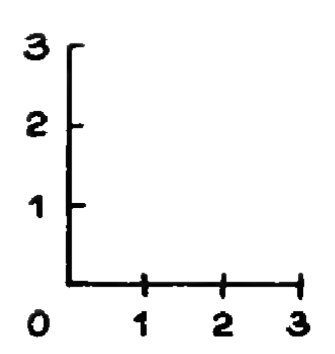


图 5

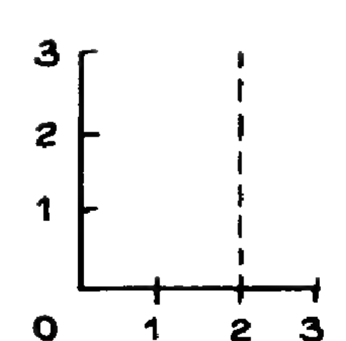
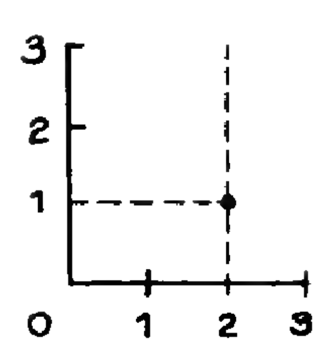


图 6



习题 B

1，绘出平面直角坐标系，并描绘下列各点：

(1,1)　　(2,2)　　(3,3)　　(4,4)

你能从中得到什么结论？

2. 下面四个点中的三个在一直线上，哪一个点不在直线上？它在直线的上方还是下方？

(0,0)　　(0.5,0.5)　　(1,2)　　(2.5,2.5)

3. 下表有四个点，每一个点的 y 坐标都是由 x 坐标按 $y=2x+1$ 计算而得。将下表填空，然后描出这四个点。你能得出什么结论？

x	y
1	3
2	5
3	—
4	—

4. 下面的图 7 有一阴影部分，(1,2)和(2,1)这两个点中哪一个在阴影中？

图 7

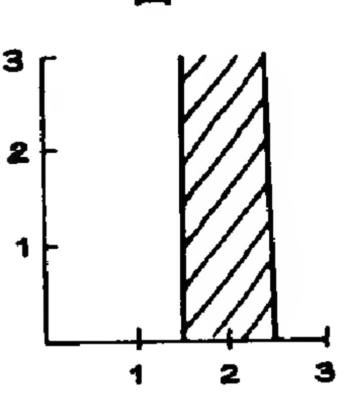


图 8

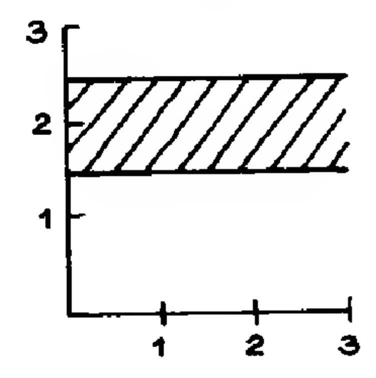
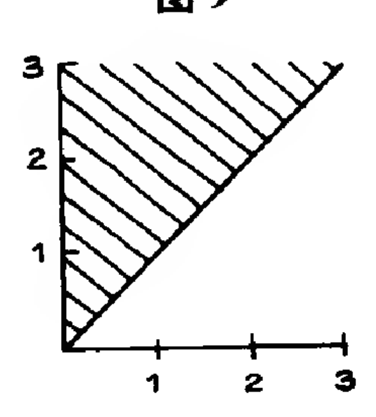


图 9



5. 见图 8，问题同上。

6. 见图 9，问题同上。

这些习题的答案在第 681 页上。



Rene Descartes（法国，1596—1650）

来自 Syracuse 大学 George Arents 研究图书馆 Woff-Leabenworth 收藏。

3. 斜率和截距

图 10 给出了一条直线。在该直线上任取一点——譬如点 A，然后沿直线向上移动到另外任一点，——譬如点 B，则 x 坐标将增加一定数量，称为前移。本例中，前移为 2。与此同时，y 坐标亦将增加一定数量，称为上升。本例中，上升为 1。注意本例中的上升是前移的一半，在该直线上不论取哪两个点，上升总是前移的一半。上升/前移的比率称为直线的斜率。

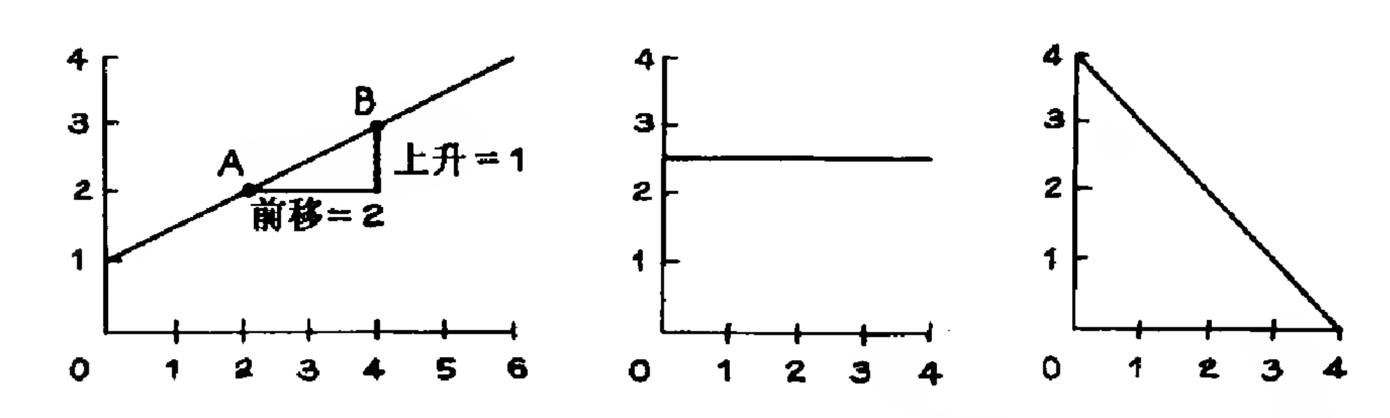
斜率＝上升/前移

斜率是沿直线随 x 的增加而 y 增加的比率。从另一个角度来理解，可以把直线想象为一条沿山而上的路，斜率测量的是路的陡峭程度。图 10 中，陡峭度为 1 比 2——对一条路来说相当陡；图 11

中，直线的斜率为 0；图 12 中，斜率为－1。若斜率为正，则直线为上升的；若斜率为 0，则直线为水平的；若斜率为负，则直线为下降的。



图 10　斜率为 1/2　　**图 11**　斜率为 0　　**图 12**　斜率为－1



直线的截距是其在 x＝0 处的高。通常情况下，坐标轴在 0 点相交，此时，截距为直线与 y 轴的交点。在图 13 中，截距为 2。有时，坐标轴并不在 0 点相交，此时，你必须多加小心。如图 14 中，坐标轴相交于(1，1)，图中直线的截距为 0—— 即该直线在 x＝0 处的高。

通常，图的坐标轴都标有单位。例如，图 15 中 x 轴的单位是英寸，y 的单位是摄氏度。斜率和截距也有相应的单位。图 15 中直线的斜率为每英寸 2.5 度，截距为－5 度。

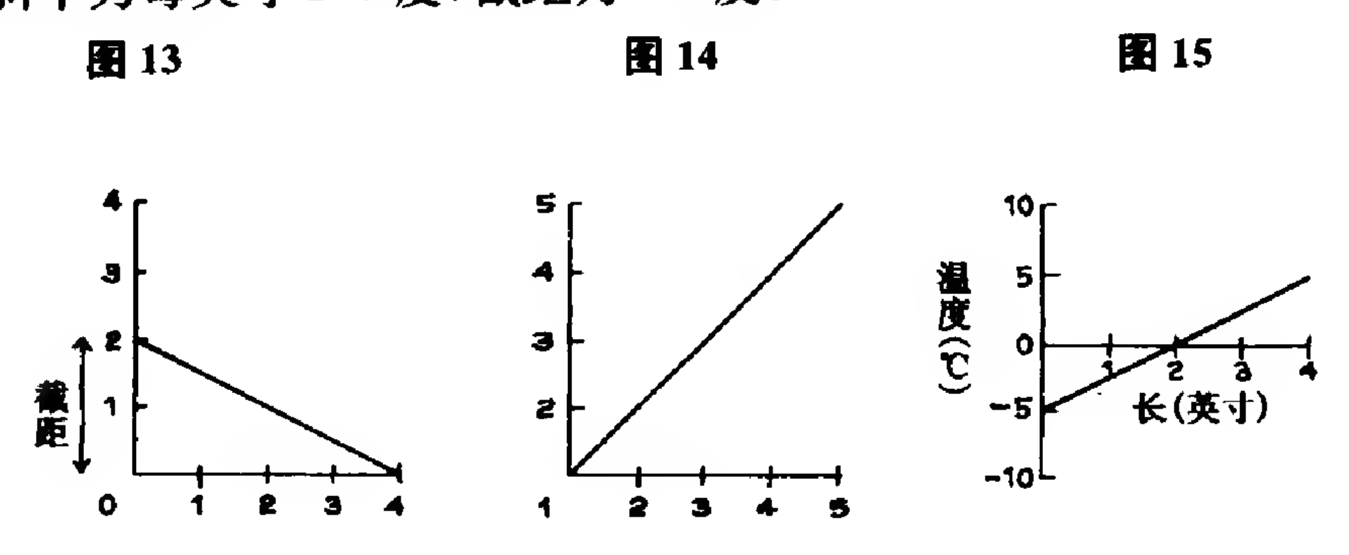
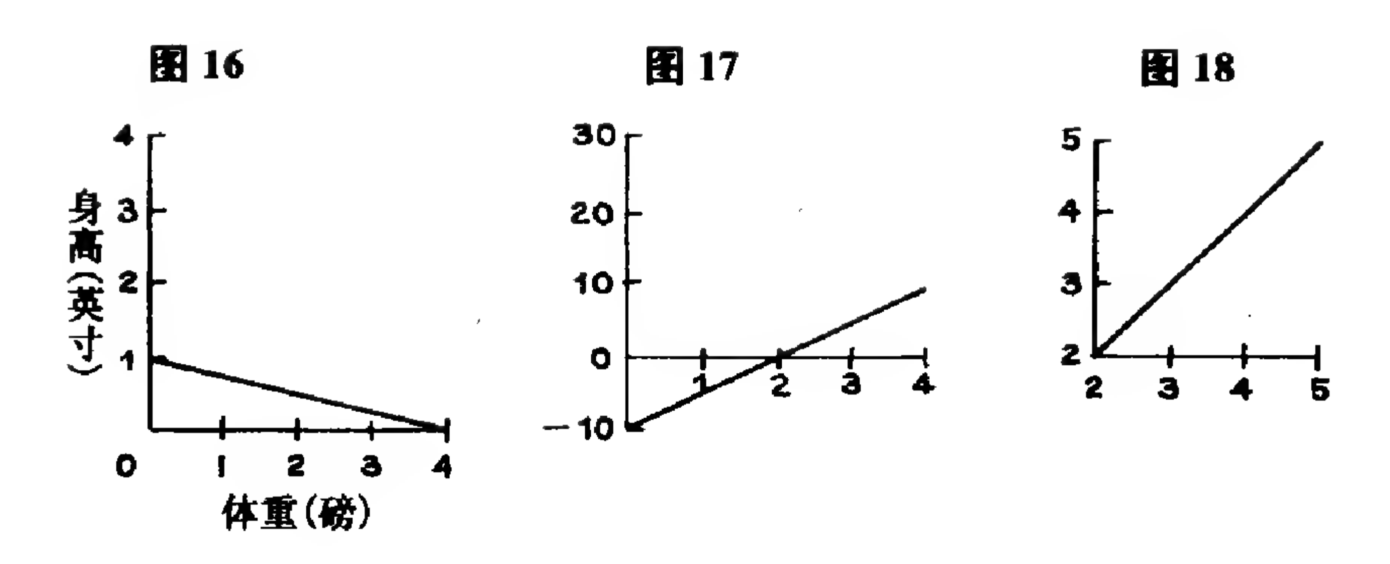


图 13　　**图 14**　　**图 15**

习题 C

1. 图 16—18 给出了直线，求出每条直线的斜率和截距。注意：坐标轴并不一定都在 0 点相交。



这些习题的答案在第 681 页上。

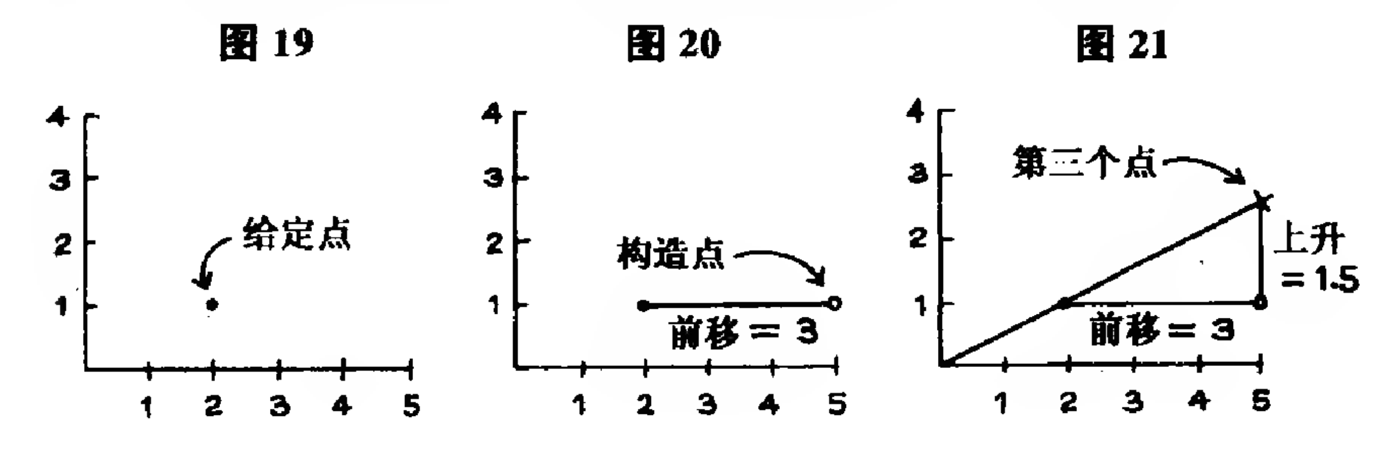
4. 描绘直线

例 1. 绘出通过点(2,1)且斜率为 1/2 的直线。

解:首先绘出一对坐标轴并描出点(2,1),如图 19 所示。然后从已知点水平向右移动任意距离:图 20 中的前移为 3,在此新位置标一构造点。由于直线是向上的,直线应在构造点的上方。在上方多高处呢?也就是说,前移为 3 时上升将为多少?此答案由斜率给出。此直线是以每一水平单位半个垂直单位的比率上升,这样,前移为 3 个水平单位时,上升为 $3\times\frac{1}{2}=1.5$。

$$上升=前移\times斜率$$

由构造点向上垂直移动 1.5,在这第三个位置标一点,如图 21 所示,这第三个点应在直线上,用直尺连接此点与已知点(2,1)即可。



习题 D

1. 绘出通过点(2,1)且有下列斜率的直线：

　　(a)　+1　　(b)　−1　　(c)　0

2. 图 21 中，若从点(2,1)出发，前移 2 上升 1，所得点是在直线上、在直线上方、还是在直线下方？
3. 问题同上，但前移 4 上升 2。
4. 问题同上，但前移 6 上升 5。
5. 绘出截距为 2，斜率为 −1 的直线。提示：此线通过点(0,2)。
6. 绘出截距为 2，斜率为 1 的直线。

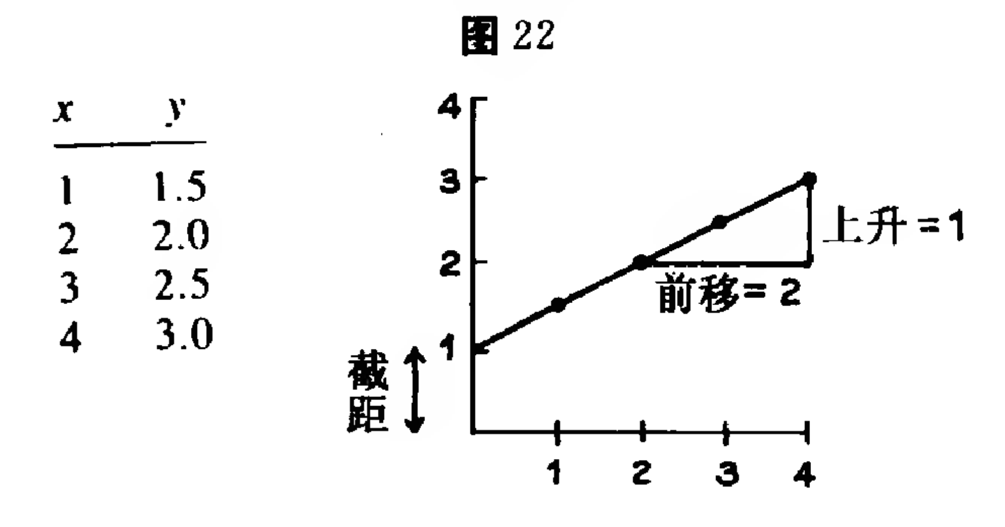
这些习题的答案在第 681—682 页上。

5. 直线的代数方程

例 2. 现有一个由点的 x 坐标计算 y 坐标的公式：$y=\frac{1}{2}x+1$。下表列出了 x 坐标为 1、2、3、4 时的点，描绘这些点，问它们是否在一条直线上？若是，求该直线的斜率和截距。

解：描点如图 22。这些点的确都在一条直线上。任何一点，只要它的 y 坐标是由 x 坐标按同一方程 $y=\frac{1}{2}x+1$ 计算得出，就都将在同一直线上。这条直线称为方程的图。该直线的斜率是 $\frac{1}{2}$，即方程中 x 的系数；截距为 1，即方程的常数项。

图 22

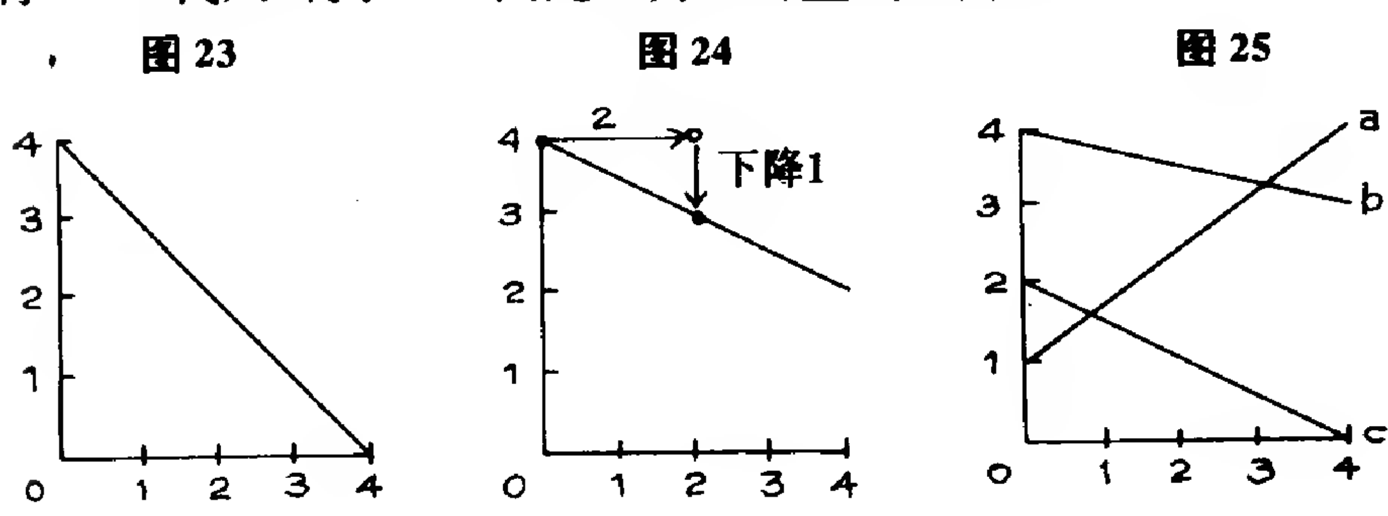


x	y
1	1.5
2	2.0
3	2.5
4	3.0

方程 $y=mx+b$ 的图是一条直线，其斜率为 m，截距为 b。

例 3. 图 23 为一条直线，求出其方程式。$x=1$ 时该直线的高为多少？

解　直线的斜率为 -1，截距为 4，故其方程式为 $y=-x+4$。将 $x=1$ 代入，得 $y=3$，因此 x 为 1 时直线的高为 3。



例 4. 绘出方程 $y=-1/2x+4$ 的直线。

解　直线的截距为 4，描点 $(0,4)$ 如图 24，该直线必然通过此点。从该点向水平方向移动任意距离，例如移动 2，由于直线的斜率为 $-1/2$，该直线将下降 1。在图 24 中标出在第一个点的右方 2、下方 1 的点，连接这两个点即得所求直线。

习题 E

1. 绘出下列方程的图：

$$(a)\ y=2x+1;\qquad (b)\ y=\frac{1}{2}x+2$$

并求出各直线的斜率和截距，以及 $x=2$ 时直线的高。

2. 图 25 中有三条直线，将它们与下列方程匹配：

$$y=\frac{3}{4}x+1\qquad y=-\frac{1}{4}x+4\qquad y=-\frac{1}{2}x+2$$

3. 描出 y 坐标是 x 坐标的 2 倍的 4 个不同的点。它们是否在一条直线上？若是，该直线的方程是什么？

4. 在同一图中描出(1,1)、(2,2)、(3,3)、(4,4)四个点。这四个点都在一条直线上,求该直线的方程。

5. 下列各点是在习题 4 中的直线上或在直线的上方,还是在直线的下方?

(a)(0,0)　　(b)(1.5,2.5)　　(c)(2.5,1.5)

6. 正确还是错误:

(a)若 y 大于 x,则点(x,y)在习题 4 中直线的上方。

(b)若 y=x,则点(x,y)在习题 4 中的直线上。

(c)若 y 小于 x,则点(x,y)在练习 4 中直线的下方。

这些习题的答案在第 682—683 页上。

第三部分　相关与回归

8
相关

有其父，必有其子。

1. 散点图

第二部分中讨论的方法对一次处理一个变量是很有效的，而要研究两个变量之间的关系，则需引入其他方法[1]。Sir Francis Galton（弗朗西斯·高尔顿爵士，英格兰 1822—1911）在研究子女与父母的相象程度时，在这方面取得了一些成果。在维多利亚时代的英格兰统计学家们受到将遗传的影响定量化这一思想的强烈吸引，并为追寻此目的收集了大量数据。我们来看一下高尔顿的学生 Karl Pearson（卡尔·皮尔逊，英格兰 1857—1936）所进行的关于家庭成员间相似性研究的成果[2]。

作为研究的一部分，皮尔逊测量了 1 078 个父亲及其成年儿子的身高。1 078 对身高数据表很难掌握，但两个变量——父亲的身高和儿子的身高——之间的关系可以从散点图中得出（见图 1）。图中每一个点代表一对父子，沿水平轴度量的点的 x 坐标给出了父亲的身高，沿垂直轴测量的点的 y 坐标给出儿子的身高。

图1　1 078 对父子身高的散点图。儿子身高等于父亲身高的家庭标绘在直线 y＝x 上，而大多数家庭散布于此线附近。散布显示了儿子身高如何不同于父亲身高。

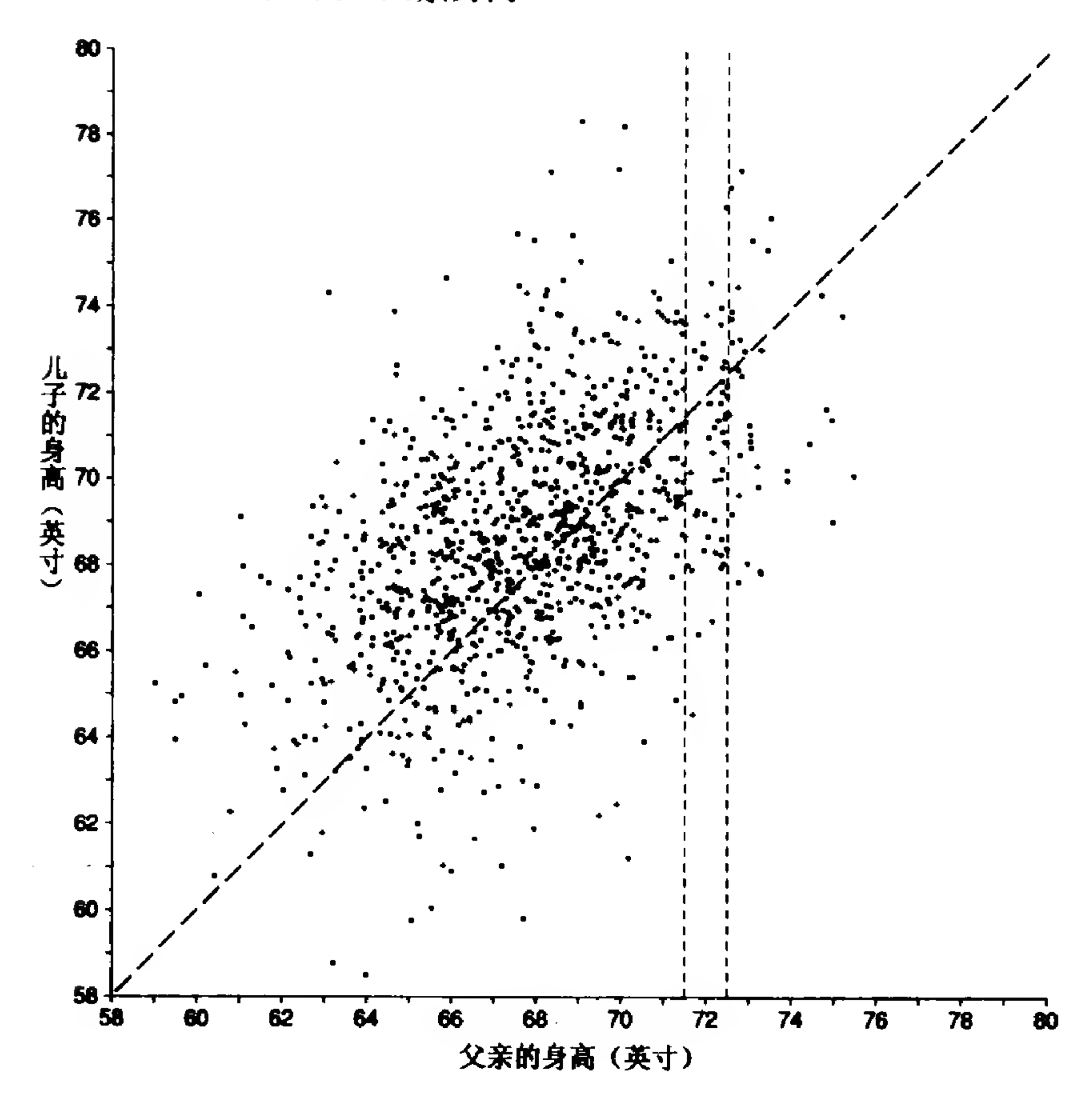
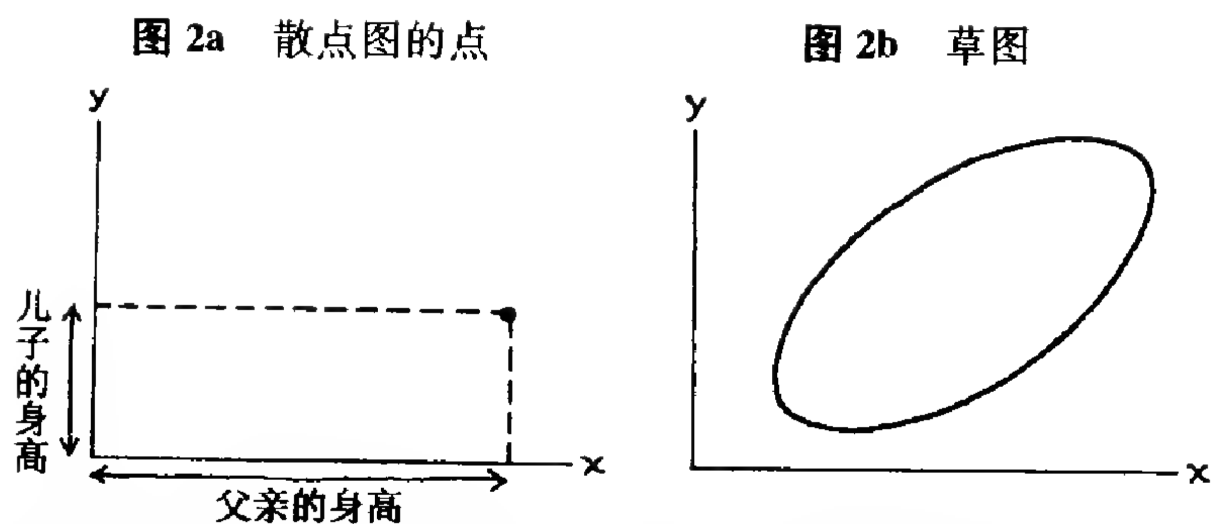


图 2a 说明了描绘散点图的原理(第 7 章有详细叙述)。图 1 的散点图象一块形如橄榄球的云，在它边缘的点很稀少。要描绘这样一个散点图的草图，只需绘出其主要的椭圆部分即可(图 2b)。

图 1 中的点群向右上方倾斜，点的 y 坐标将随 x 坐标的增加而增加。用统计学家的话说，父亲和儿子的身高之间存在着一个正相关。用日常生活中的语言来表述，一般规律，较高的父亲有较高的儿子，这符合常识。但散点图表示了这种相关是多么弱。观察图



1 的 45 度线，它对应那些儿子的身高与父亲身高相等的家庭。例如沿该线，如果父亲的身高为 72 英寸则儿子也是 72 英寸高；如果父亲的身高是 64 英寸则儿子也同样高；如此等等。类似地，如果儿子的身高与父亲相近，则散点图中的相应点将靠近此线，如图 3 中的点。

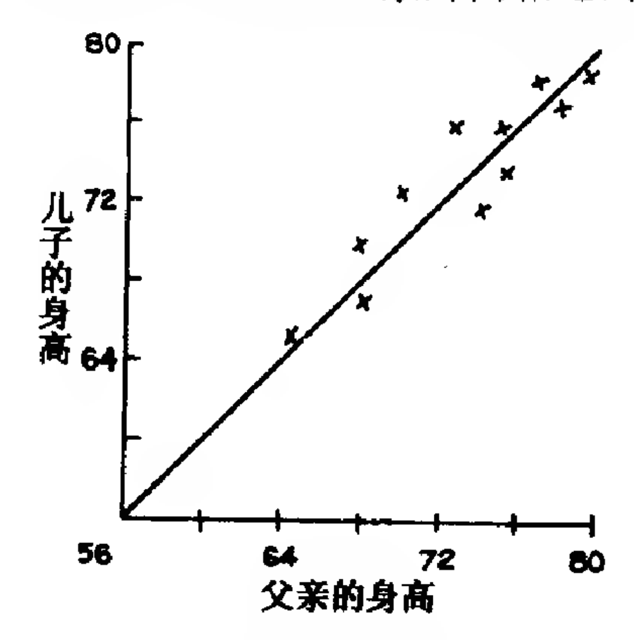


图 3 儿子身高与父亲身高相近的点。

然而，实际散点图中在 45 度线的周围有着比图 3 更多的散布，这种散布说明父亲的身高与儿子的身高之间的相关关系是很弱的。为了理解散布的含义，假定你必须猜测儿子的身高：父亲的身高能给你多大帮助呢？例如，假定父亲是 72 英寸高，图 1 中烟囱状内的点代表了所有父亲身高近似到英寸为 72 英寸的父子对，换言之，此时父亲的身高是在 71.5 英寸和 72.5 英寸之间（即图中两条虚纵线与 x 轴相交处），如烟囱状内纵向分布的点所示，儿子的身高仍有许多变化。因此，即便你碰巧知道父亲的身高，在试图预

测他的儿子的身高中误差仍有大量余地。

> 如果两个变量之间存在强相关，则已知一个变量的值对预测另一个变量的值将很有帮助。但若是弱相关，关于一个变量的信息对猜测另一个变量的值无多大帮助。

社会科学中研究两个变量之间的关系时，常常把一个变量称为自变量，另一个变量称为因变量。通常认为是自变量影响因变量，而不是因变量影响自变量。图 1 中，父亲的身高取作自变量，并沿 x 轴描绘：父亲的身高影响儿子的身高。然而，没有什么东西可以阻止研究者把儿子的身高作为自变量，这种选择也许是恰当的。例如，如果问题是从儿子的身高猜测父亲的身高。



Sir Francis Galton（弗朗西斯·高尔顿爵士，英格兰，1822—1911）
摘自：Biometrika(1903)。

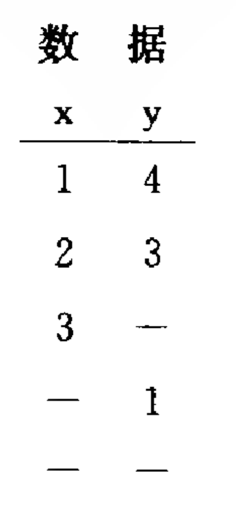
在进一步学习之前，最好先完成本节的习题。这些习题很简

单，它们将真正地帮助你理解本章的其他内容。完成习题有困难时，可参阅第七章。

练习 A

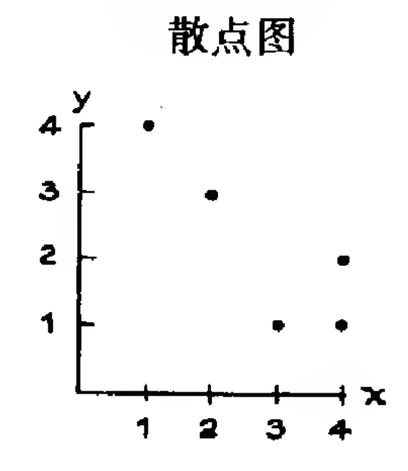
1. 下面给出一些虚拟数据且一并给出散点图，填表中的空白。

数　据



x	y
1	4
2	3
3	—
—	1
—	—

散点图



2. 下面是一些虚拟数据的散点图：

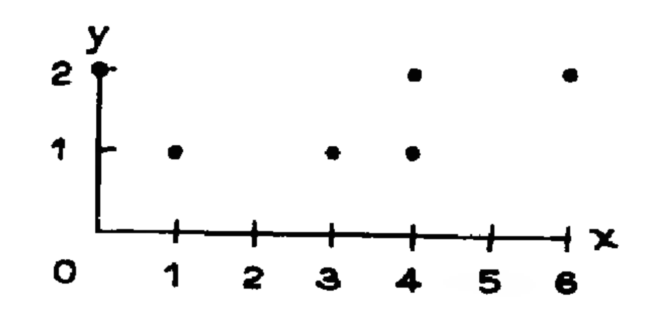
(a)x 的平均数大约是 1、1.5 还是 3?

(b)y 的平均数大约是 1、1.5 还是 3?

(c)x 的诸值和 y 的诸值，哪一个更分散?

(d)写出该散点图如习题 1 中那样的数据表。

(e)计算诸 x 值的平均数和 SD 及诸 y 值的平均数和 SD，以检验(a)、(b)、(c)。



3. 下图中：

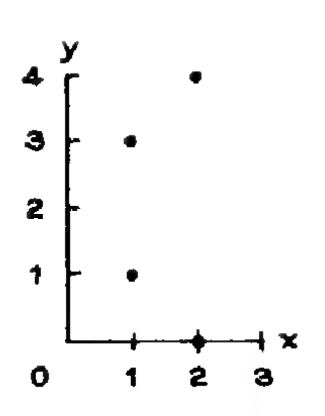
(a)诸 x 值的平均数大约是 1、1.5 还是 2?

(b)诸 x 值的 SD 大约是 0.1、0.5 还是 1?

(c)诸 y 值的平均数大约是 1、1.5 还是 2?

(d)诸 y 值的 SD 大约是 0.5、1.5 还是 3?

4. 对下列每一组虚拟数据描绘散点图。标为“x”的变量沿 x 轴描绘，标为“y”的变量沿 y 轴描绘。完整地标出每一根坐标轴。有的情况下你必须不止一



次地描绘同一个点，为方便起见，重复描绘同一点的次数可标在该点旁的小圆圈内，如下图所示。

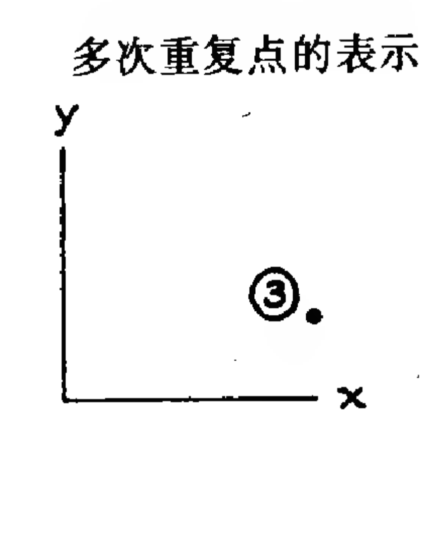
(a)

x	y
1	2
3	1
2	3
1	2

(b)

x	y
3	5
1	4
3	1
2	3
1	4
4	1

多次重复点的表示



5. 用图 1 回答下列问题：

(a)最矮的父亲的身高是多少？他的儿子的身高是多少？

(b)最高的父亲的身高是多少？他的儿子的身高是多少？

(c)取所有父亲的身高为 72 英寸(近似到英寸)的家庭，最高的儿子身高是多少？最矮的儿子身高是多少？

(d)儿子约为 77 英寸高的家庭有多少？他们的父亲的身高为多少？

(e)父亲的平均身高约为

64 英寸　　68 英寸　　72 英寸

(f)父亲身高的 SD 约为

3 英寸　　6 英寸　　9 英寸

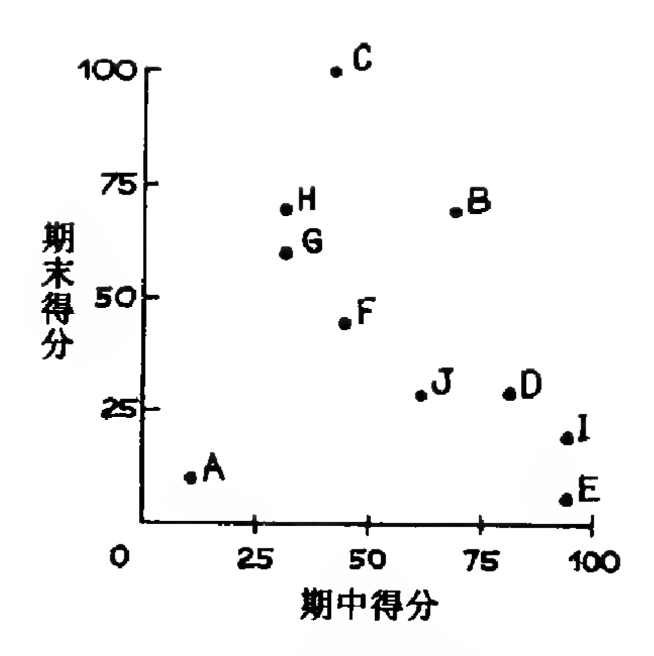
6. 各为 A、B、C、D、E、F、G、H 和 I 的学生参加了一门课程的期中和期末考试，得分的散点图如下：

(a)哪些学生的期中和期末考试成绩相等？

(b)哪些学生的期末成绩比期中高？

(c)期末平均成绩约为 25、50 还是 75？

(d)期末成绩的 SD 约为 10、25 还是 50？

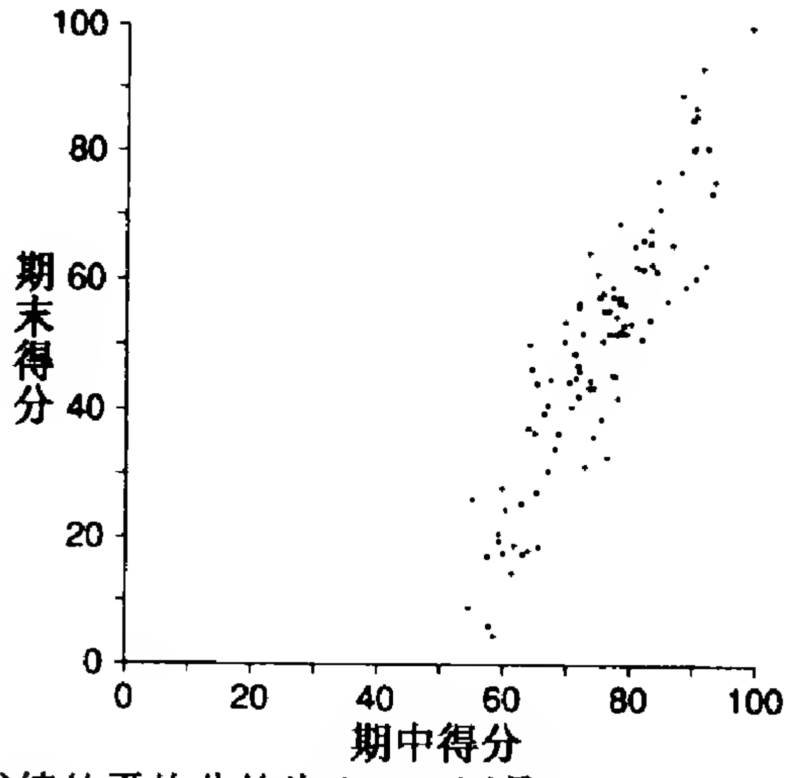


(e)期中考试成绩在 50 分以上的学生的期末考试成绩平均约为 30、50 还是 70?

(f)正确还是错误:总的来说期中考试成绩好的学生在期末考试中也好。

(g)正确还是错误:在期中和期末考试之间存在着一个强正相关。

7. 下面的散点图给出了某一课程的期中考试和期末考试成绩:



(a)期中考试成绩的平均分约为 25、50 还是 75?

(b)期中考试成绩的 SD 约为 5、10 还是 20?

(c)期末考试成绩的 SD 约为 5、10 还是 20?

(d)期中考试和期末考试哪一个更难?

(e)哪一次考试的分数较为离散?

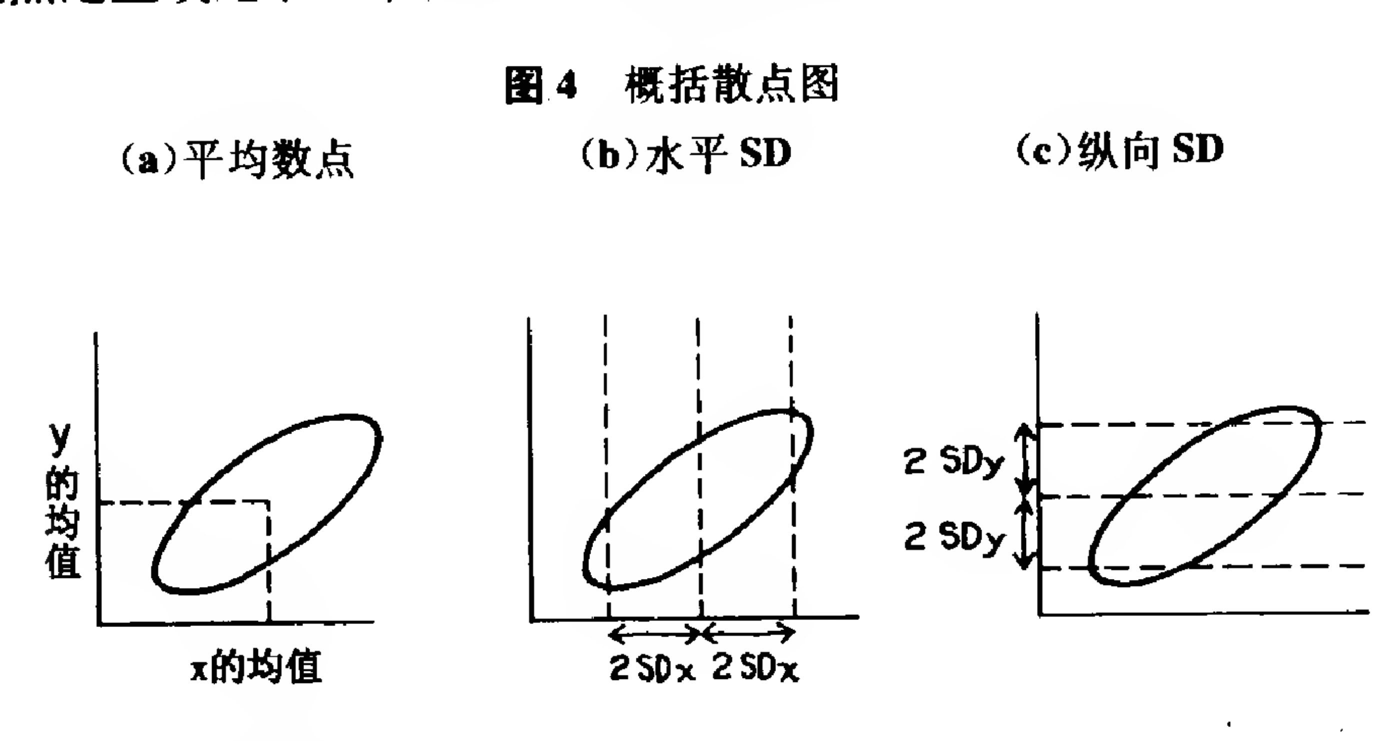
(f)正确还是错误:期中和期末考试分数间存在强正相关关系。

这些习题的答案在第 684—685 页上。

2. 相关系数

假定你正在考察两个变量间的关系并已描绘了散点图。此图形如点云。如何对图进行数量概括呢？第一步应标出表示 x 的平均数和 y 的平均数的点(图 4a)，此点称为平均数点，[3]它确定了点云的中心。下一步应度量点云从这边到那边的散布度。这可通过诸 x 值的 SD——水平 SD——得到。大多数点都将落在平均数点两侧 2 个水平 SD 的范围内(图 4b)。同样，诸 y 值的 SD——纵向 SD——也可用来度量从上到下的散布度。大多数点都将落在平均数点之上或之下 2 个纵向 SD 的范围内(图 4c)。

图 4　概括散点图



至此，概括统计量是：

· 诸 x 值的平均数，诸 x 值的 SD；

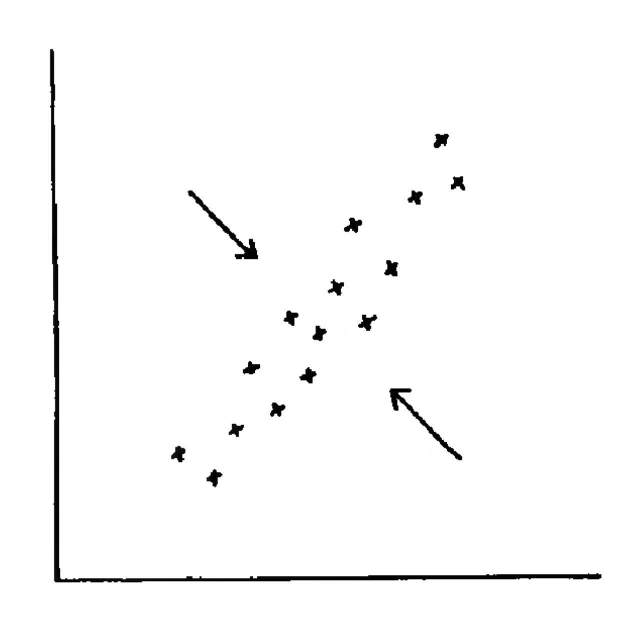
· 诸 y 值的平均数，诸 y 值的 SD。

这些统计量告诉我们点云的中心，及点云的横向和纵向散布度，但还缺一个东西——两个变量之间相关的程度。观察图 5 的散点图。两块点云有相同的中心及同样的横向、纵向散布度。然而，第一个图中的点紧密地群集在一条直线周围：这两个变量间存在一个强的线性相关关系。在第二个图中，点的群集要松散得多。两个图中相关的程度是不一样的。为了度量相关程度，需引入另一个

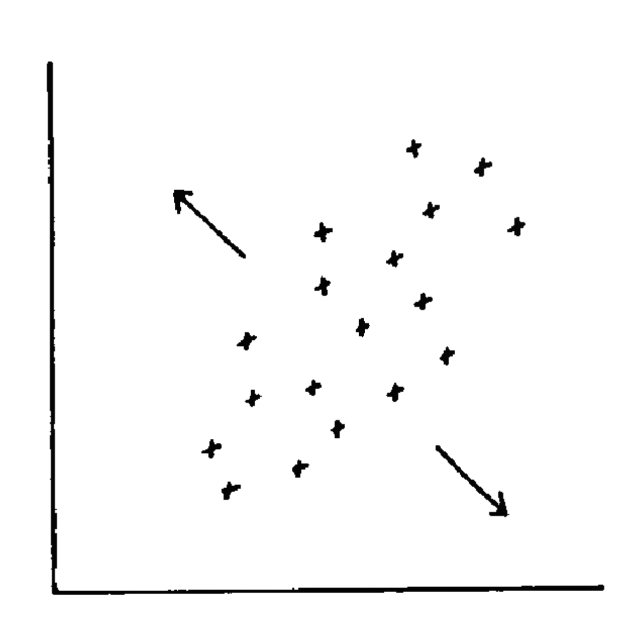
概括性统计量——相关系数。相关系数通常缩写为 r，这是约定俗成的记号（与“Correlation”中有两个 r 无关）。

图 5　概括散点图。点云关于直线相关系数的测定

(a)相关程度接近 1，点云紧密集中在直线周围



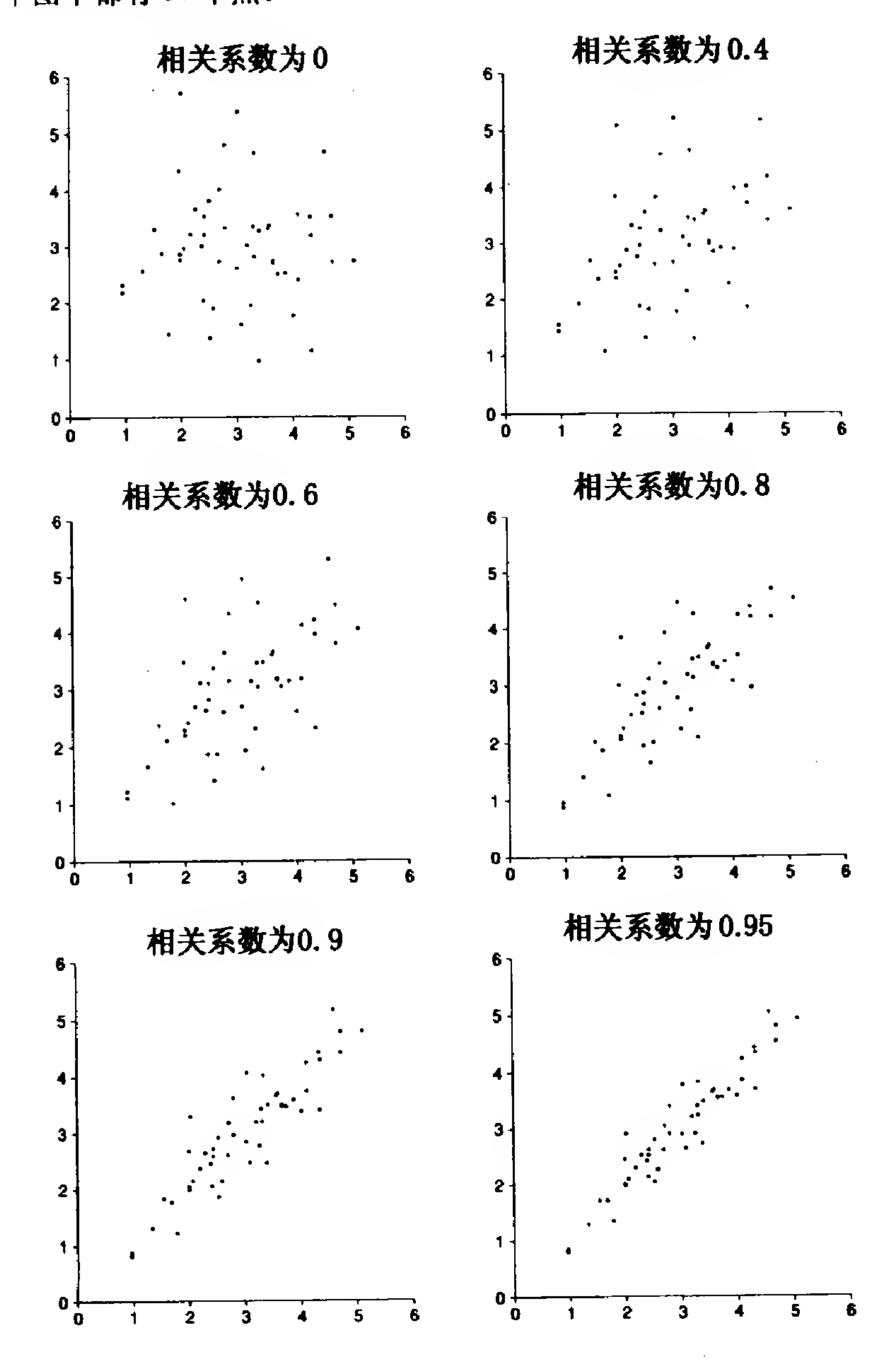
(b)相关程度接近 0，点云很松散



相关系数是线性相关或围绕直线群集程度的一种度量。两个变量间的关系可用如下统计量进行概括： • 诸 x 值的平均数，诸 x 值的 SD • 诸 y 值的平均数，诸 y 值的 SD • 相关系数 r

r 的计算公式将在第四节介绍，目前我们重点进行图形解释。图 6 给出有关六个虚拟数据的散点图，每一个都有 50 个点。这些图都是用计算机产生的。在所有六个图中，x 和 y 的平均数都是 3，SD 都是 1。计算机在图的上方打印出了各个图的相关系数。左上图的相关系数为 0，图中的点是杂乱无章的，当 x 增加时，y 没有增加或减少的趋势，而是无规律地变化。

图 6 相关系数——6 个正相关的情形。图中横向和纵向的平均数都是 3，SD 都为 1。点围绕直线的群集程度用相关系数度量。每个图中都有 50 个点。



第二个散点图的 r=0.40；线性关系开始显现。第三个图的 r=0.60，有较强的线性关系。如此直到最后一个图。r 越接近于 1，变量间的线性关系就越强，点在直线周围的群集就越紧密。等于 1 的相关系数，这种情形未在图中出现，通常称为完全相关，表示所有点都位于一条直线上，因此在变量间存在着一个完全的线性关系。相关系数总是等于或小于 1 的。

例如，同卵双胞胎身高间的相关系数是 0.95，[4]这些双胞胎是否有相同的身高呢？若是，他们在散点图中的点将紧靠直线 y=x。图 6 中右下图的相关系数是 0.95，但是在直线周围仍有相当量的散布。双胞胎的散点图与此类似，身高远不是完全相同的。

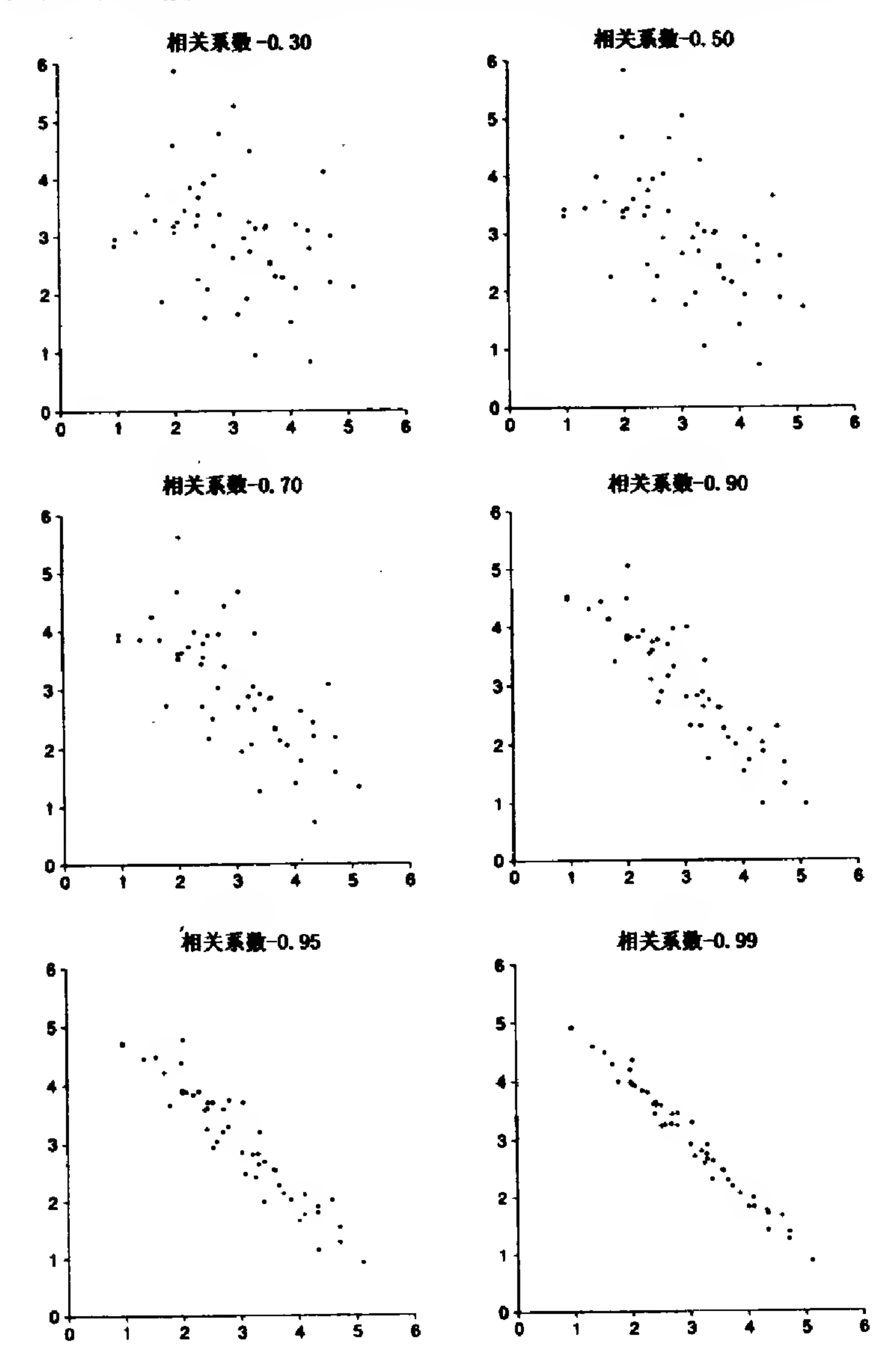
又如，1988 年美国 25－34 岁男子的收入与受教育程度间的相关系数是 0.34，55－64 岁男子的相关系数则上升为 0.44。[5]如图 6 的散点图所显示的，收入与受教育程度间的关系对年老一些的人要强一些，但这仍很粗略。弱相关是社会科学研究中的通例，r 的值通常都在 0.3 至 0.7 之间。

要注意的是：r=0.80 并不意味着 80%的点都紧密地群集在一条直线的周围，也不表示其线性程度是 r=0.40 情况时的二倍。就目前来说，尚无直接的方式去解释相关系数的确切数值；这一工作将在第 10 和 11 章中进行。

到目前为止，我们仅讨论了正相关。在美国，妇女受教育的程度越高，所生的孩子数越少，这是一种负相关现象：总的说来，受教育程度的增加将伴随着孩子数的减少。负相关可被相关系数前有负号所显示。图 7 给出了另外六组虚拟数据的散点图，每个图中都有 50 个点，如图 6 一样，每一个变量的平均数都是 3，SD 都是 1。

例如，－0.90 的相关系数表示了如 0.90 相关系数同样的群集程度。相关系数为负时点沿斜向下的直线群集，为正时点沿斜向上的直线群集。1988 年 25－34 岁的妇女受教育程度与其子女数间的相关系数为－0.2，是弱负相关[6]。相关系数为－1 是完全负相关，所有点都将在一条斜向下的直线上。

图 7　相关系数——六个负值。图中横向和纵向平均数都为 3，SD 都为 1。点沿直线群集程度由相关系数测定。每个图都有 50 个点。



相关系数总是在－1和1之间，但可以是此范围内的任何值。正相关时散点云是斜向上的，此时一个变量增加，另一个变量也增加；负相关时散点云是斜向下的，此时一个变量增加，另一个变量将减少。

习题 B

1. (a)二手车车龄与其价格之间的相关是正还是负？为什么？(作古玩的车除外。)

   (b)车重与每加仑油所跑英里数间的相关怎么样呢？

2. 下面的散点图中

   (a)x 的平均数约为

   1.0　1.5　2.0　2.5　3.0　3.5　4.0

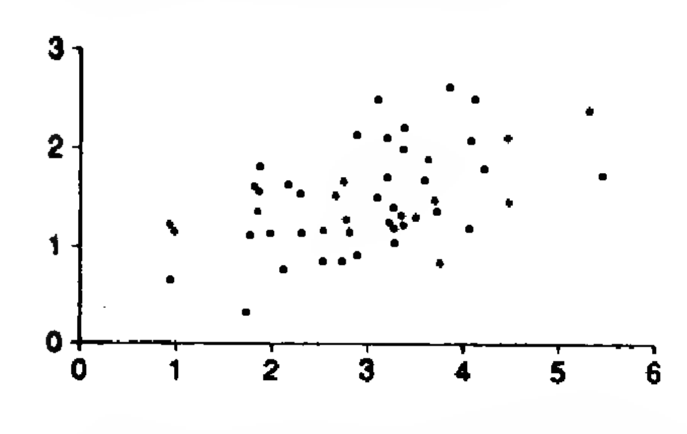
   (b)对 y 的平均数，问题同(a)。

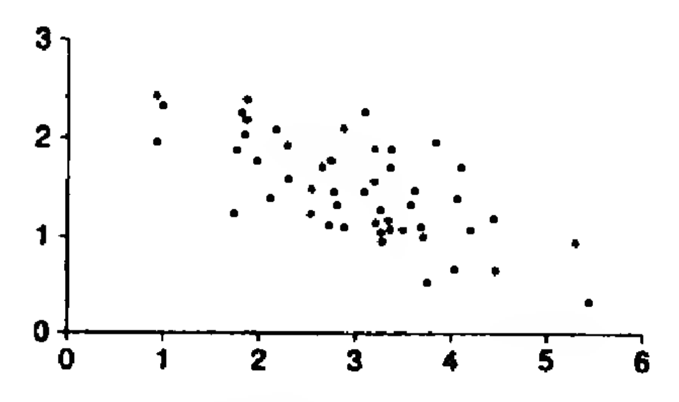
   (c)x 的 SD 约为

   0.25　0.5　1.0　1.5

   (d)对 y 的 SD，问题同(c)。

   (e)相关是正、是负、还是 0？





3. 上题中哪个散点图的相关系数更接近 0(不考虑正负号)？

4. 在图 1 中，父亲和儿子身高间的相关系数约为－0.3、0、0.5 还是 0.8？

5. 在图 1 中，如果仅考虑身高在 6 英尺以上的父亲和他们的儿子，身高间的相关系数约为－0.3、0、0.5 还是 0.8？

6. 如果妇女总是嫁给比她们大 5 岁的男子，夫妇年龄间的相关系数是多少？为什么？

7. 美国丈夫和妻子年龄之间的相关系数是________。取一个选择，并解释。

恰为－1，接近－1，接近 0，接近 1，恰为 1

8. 调研人员正对在加利福尼亚大学注册的学生的情况进行调查。学生要在问卷中填写他们的出生年份、年龄(岁)、母亲年龄，等等。在下列备选答案中选择填空并解释。

(a)学生年龄与出生年份间的相关系数为______。

(b)学生年龄与母亲年龄间的相关系数为______。

－1　　接近－1　　某负数

0　　某正数　　接近 1　　1

9. 正确还是错误并解释之：如果相关系数是 0.90，则 90％的点高度相关。

这些习题的答案在第 685 页上。

3. SD线

在象图 6 的一系列散点图中，当 r 越接近 1 时，点将越密集地群集在一条直线旁。这是条什么样的直线呢？我们把它称为SD线，它穿过对两个变量来说与平均数的差都为 SD 的同样倍数的所有点。现以身高与体重的散点图为例说明，若某人的身高恰比平均身高多 1SD，体重也比平均体重多 1SD，则他的点将落在 SD 线上。而身高比平均身高多 1SD，但体重比平均体重多 0.5SD 的人不在此线上。同样，身高在平均身高之下 2SD，体重也在平均体重之下 2SD 的人在此线上。身高在平均身高之下 2SD，但体重在平均体重之下 2.5SD 的人不在此线上。

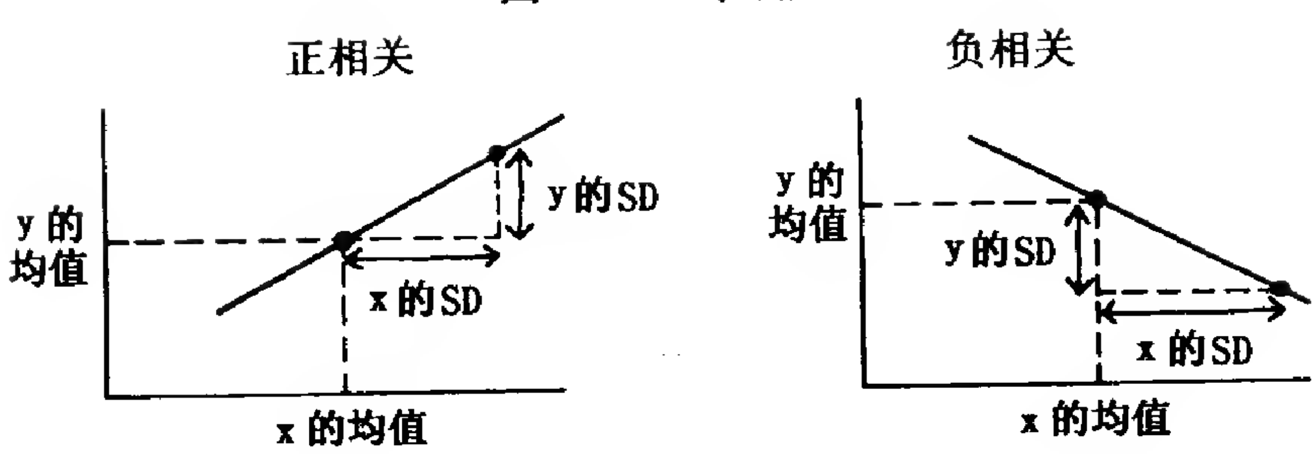
图 8 说明了如何在图上描绘 SD 线。该线穿过平均数点，并以横向每增加 1SD，纵向也增加 1SD 的比例上升。用专门的术语表达，斜率是比

(y 的 SD)/(x 的 SD)

这在正相关时适用。当相关系数为负时，SD 线向下降；斜率为[7]

－(y 的 SD)/(x 的 SD)

图 8　SD 线的描绘



习题 C

1. 正确还是错误：

 (a)SD 线穿过平均数点。

 (b)SD 线穿过(0,0)点。

2. 一项关于男大学生的研究发现。他们的平均身高为 69 英寸，SD 为 3 英寸，平均体重为 140 磅，SD 为 20 磅。相关系数是 0.60。如果他们中某人为 72 英寸高，其体重应为多少才能使他的点在 SD 线上？

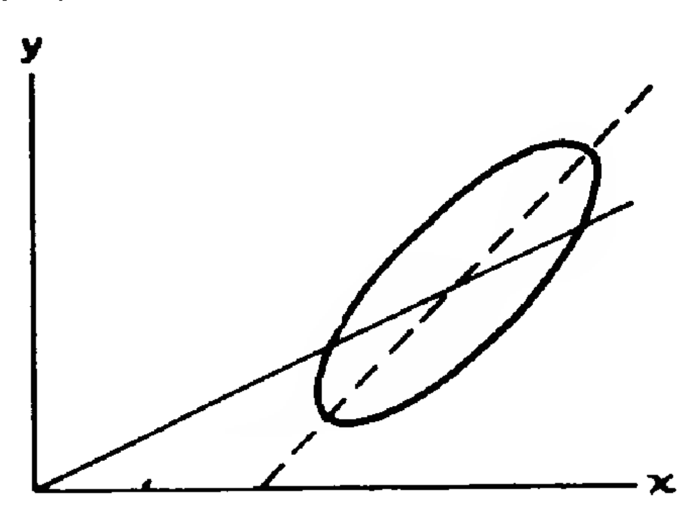
3. 数据同习题 2，指出下列每一个学生是否在 SD 线上：

 (a)身高 75 英寸，体重 180 磅

 (b)身高 66 英寸，体重 130 磅

 (c)身高 66 英寸，体重 120 磅

4. 在下面的散点图中，说出实线和虚线哪一条是 SD 线？



这些习题的答案在第 686 页上。

4. 计算相关系数

相关系数的计算过程可叙述如下：

将每个变量都转换为标准单位。乘积的平均数即为相关系数。

将第一个变量记为 x，第二个变量记为 y，相关系数记为 r，则可用公式表示如下：

r＝[（以标准单位表示的 x）×（以标准单位表示的 y）]的平均数

例 1. 计算表 1 中虚拟数据的 r。

表 1　数据

x	y
1	5
3	9
4	7
5	1
7	13

注：表 1 第一行代表研究对象的两个测量值；两个数字分别为散点图中对应点的 x 和 y 坐标。其他行亦如此。

解：计算过程如表 2 所示。

第一步：按第 5 章所述方法将 x 值转换为标准单位，这一步的计算量较大。首先，应求出诸 x 值的平均数和 SD：

表 2　r 的计算

x	y	以标准单位表示的 x	以标准单位表示的 y	乘积
1	5	－1.5	－0.5	0.75
3	9	－0.5	0.5	－0.25
4	7	0.0	0.0	0.00
5	1	0.5	－1.5	－0.75
7	13	1.5	1.5	2.25

x 的平均数=4，SD=2

然后，将每一 x 值减去平均数，并除以 SD：

$$\frac{1-4}{2}=-1.5 \quad \frac{3-4}{2}=-0.5 \quad \frac{4-4}{2}=0 \quad \frac{5-4}{2}=0.5 \quad \frac{7-4}{2}=1.5$$

此结果见表 2 的第三栏。这些数字告诉你以 SD 为单位，x 值在平均数之上或之下多远。例如，第一个值在平均数之下 1.5SD。

第二步：将 y 值转换为标准单位，结果见表 2 的第 4 栏。这就完成了计算中最繁复的部分。

第三步：对表中各行，求如下乘积：

（以标准单位表示的 x）×（以标准单位表示的 y）

结果见表 2 最后一栏。

第四步：求出这些乘积的平均数：

r=（以标准单位表示的 x）×（以标准单位表示的 y）的平均数

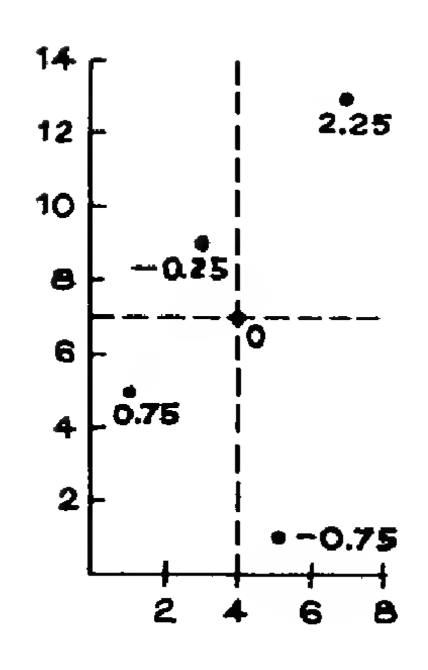
$$=\frac{0.75-0.25+0.00-0.75+2.25}{5}=0.40$$

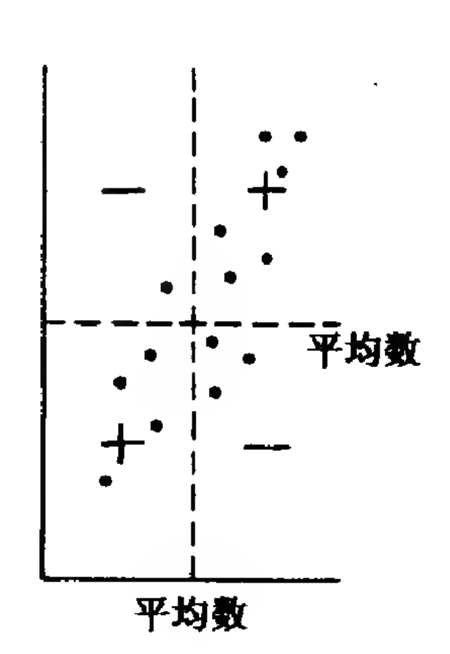
解毕。如果对该数据绘出散点图（见图 9a），散点将斜上升并松散地群集于直线旁。

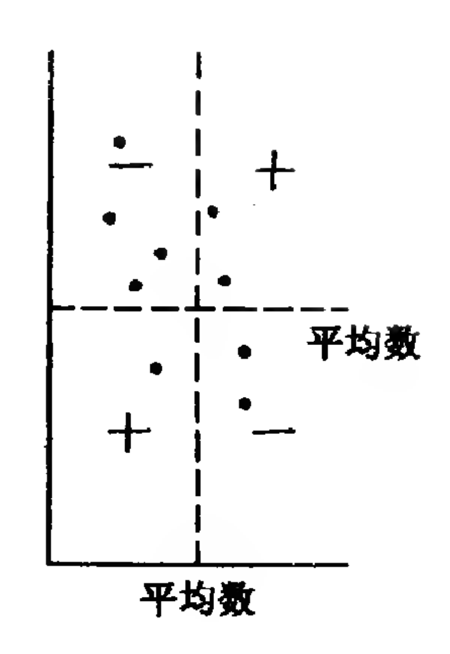
为什么 r 能度量相关程度呢？在图 9a 中，各点旁都标明了相应的乘积值。过平均数点绘出水平线和垂直线，它们将散点图分为四个象限。如果一个点在左下象限，两个变量都小于平均数因而以标准单位是负的，两个负数之积为正的。对右上象限，两个正数之积仍为正。对其余两个象限，一正一负之积为负。所有这些乘积之平均数即为相关系数。若 r 为正，则两个正象限中的点占主导地位，如图 9b 所示；若 r 为负，则两个负象限中的点起主导作用，如图 9c 所示。

图 9　相关系数的原理

(a)表 1 的散点图　　　　(b)正 r　　　　(c)负 r







习题 D

1. 计算下列表中数据的 r

(a)

x	y
1	6
2	7
3	5
4	4
5	3
6	1
7	2

(b)

x	y
1	2
2	1
3	4
4	3
5	7
6	5
7	6

(c)

x	y
1	7
2	6
3	5
4	4
5	3
6	2
7	1

2. 找出图 6 中相关系数为 0.95 的散点图。图中两个变量同时都在平均数之上的点所占百分比约为

　　5%　　25%　　50%　　75%　　95%

3. 对相关系数为 0.00，重复习题 2。

4. 利用图 7，取相关系数为－0.95，重做习题 2。

这些习题的答案在第 686 页上。

技术性注：有时可用另一种计算相关系数的方法[8]：

$$r = \frac{\mathrm{cov}(x,y)}{(x\text{ 的 SD})\times(y\text{ 的 SD})}$$

其中 cov(x,y)=(乘积 xy 的平均数)−(x 的平均数)×(y 的平均数)

5. 复习题

复习题可能包含前面各章的内容。

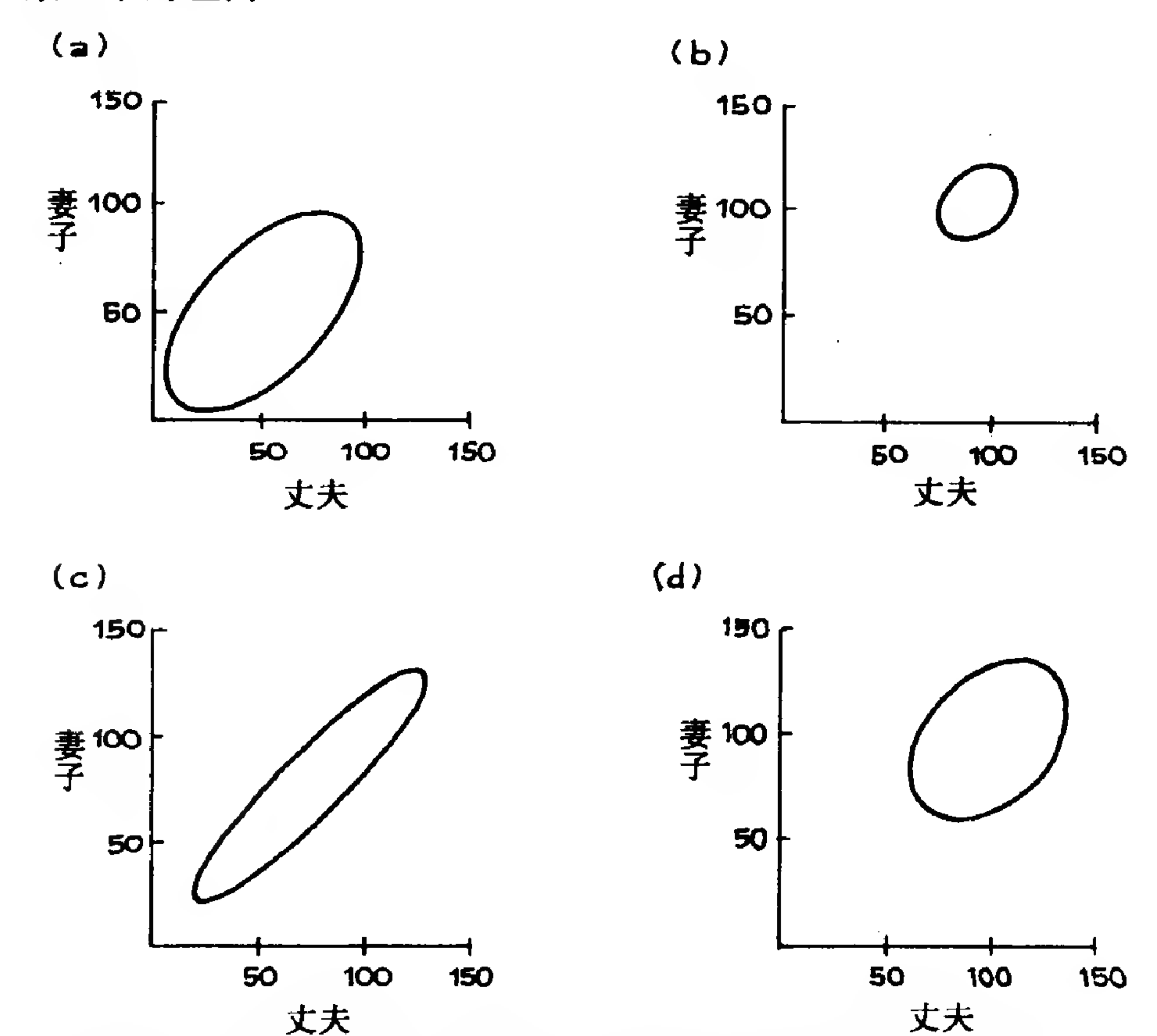
1. 一项关于丈夫和妻子智商的研究得到了如下结果：

丈夫：平均智商=100，SD=15

妻子：平均智商=100，SD=15

$$r=0.6$$

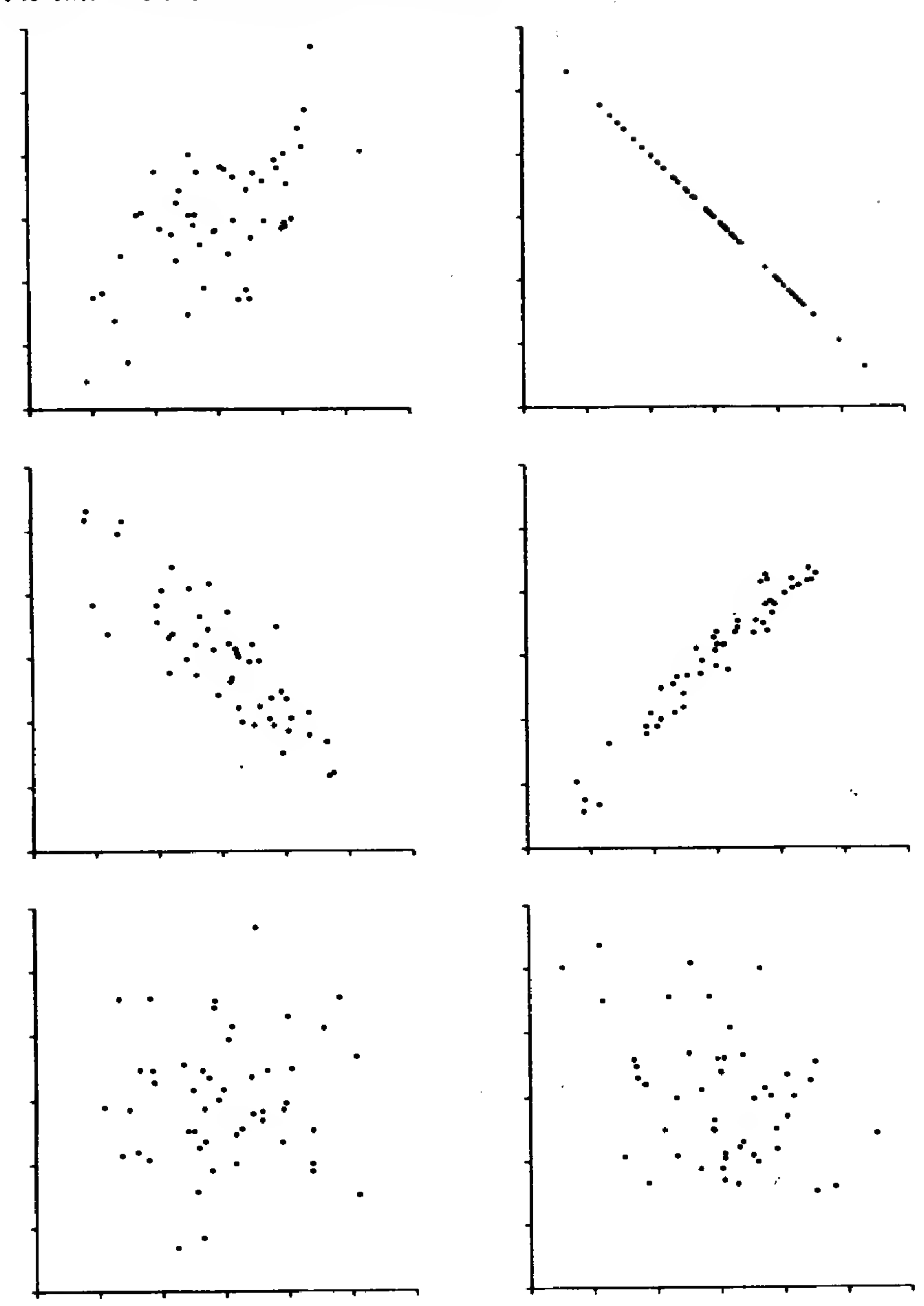
下面的散点图之一与此数据对应，是哪一个？简要说明否定其余几个的理由。



2. 下面是六个虚拟数据的散点图，已被打乱顺序的相关系数为

−0.85　　−0.38　　−1.00　　0.06　　0.97　　0.62

将散点图与相关系数配伍(见下图)。



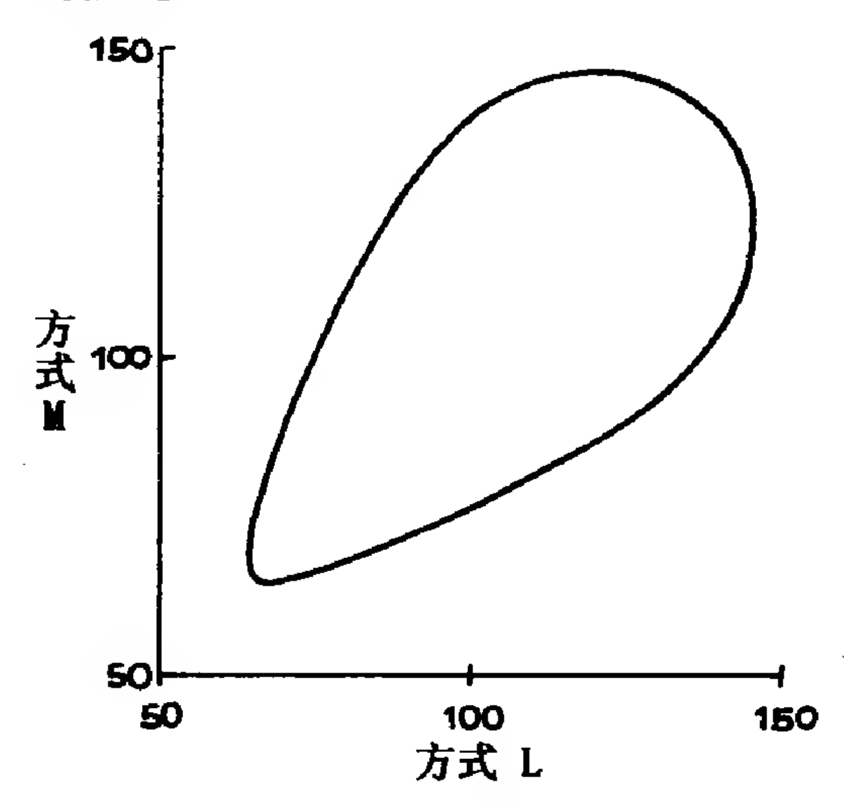
3. 在一项大型的心理学研究中,每一测试对象都参加了两次智商测验(用 Stanford－Binet 的 L 方式和 M 方式),测试结果的散

点图已绘制如下。现在试图要从L方式的测试分数预测M方式的测试分数，下述结论哪些是正确的，为什么？

(i)当L方式的分数为75时预测较容易。

(ii)当L方式的分数为125时预测较容易。

(iii)(i)和(ii)差不多。



4. 现已收集了三组数据，各自的相关系数也已算出，变量是

(i)学生在第一学年和第二学年的平均等级分。

(ii)学生在第一学年和第四学年的平均等级分。

(iii)2×4的木板的长和重量。

打乱了顺序的相关系数为：

0.30　　0.60　　0.95

将相关系数与数据配伍，并说明理由。

5. (a)对汽车的代表性样本，车龄与汽油消耗量(每加仑英里)间的相关系数是正还是负？

(b)汽油消耗量与车主收入间的相关系数为正[9]，如何解释这个正相关？

6. 假若男人总是要比他们正好矮8%的女子为妻，他们身高间的相关系数是多少？

7. 美国的丈夫和妻子身高间的相关系数大约为−0.9、−0.3、0.3还是0.9? 解释之。

8. 某助教在班上进行了一次测验,共有10道题且不给出各题的分数。评卷后,该助教记下了每个学生答对和答错的题数。正确答案的平均数为6.4,SD为2.0,错误答案的平均数为3.6,SD同样是2.0。相关系数为

   0　−0.50　+0.50　−1　+1　没有数据不能确定

   从中选择一个答案并给予解释。

9. 求出下列三组数据的相关系数。

(a) x	(a) y	(b) x	(b) y	(c) x	(c) y
1	5	1	1	1	2
1	3	1	2	1	2
1	5	1	1	1	2
1	7	1	3	1	2
2	3	2	1	2	4
2	3	2	4	2	4
2	1	2	1	2	4
3	1	3	2	3	6
3	1	3	2	3	6
4	1	4	3	4	6

10. 某班期末考试和期中考试分数间的相关系数为0.50,期末考试和家庭作业分数间的相关系数为0.25。正确还是错误并解释之:期末考试和期中考试分数间线性相关的程度是期末考试和家庭作业分数间线性相关程度的二倍。

11. 对人的生长情况的跟踪研究自1929年就在柏克利人类发展研究所进行[⑩],下面的散点图给出了64个男孩子在4岁和18岁时的身高。

   (a)4岁时的平均身高约为

   38英寸　　42英寸　　44英寸

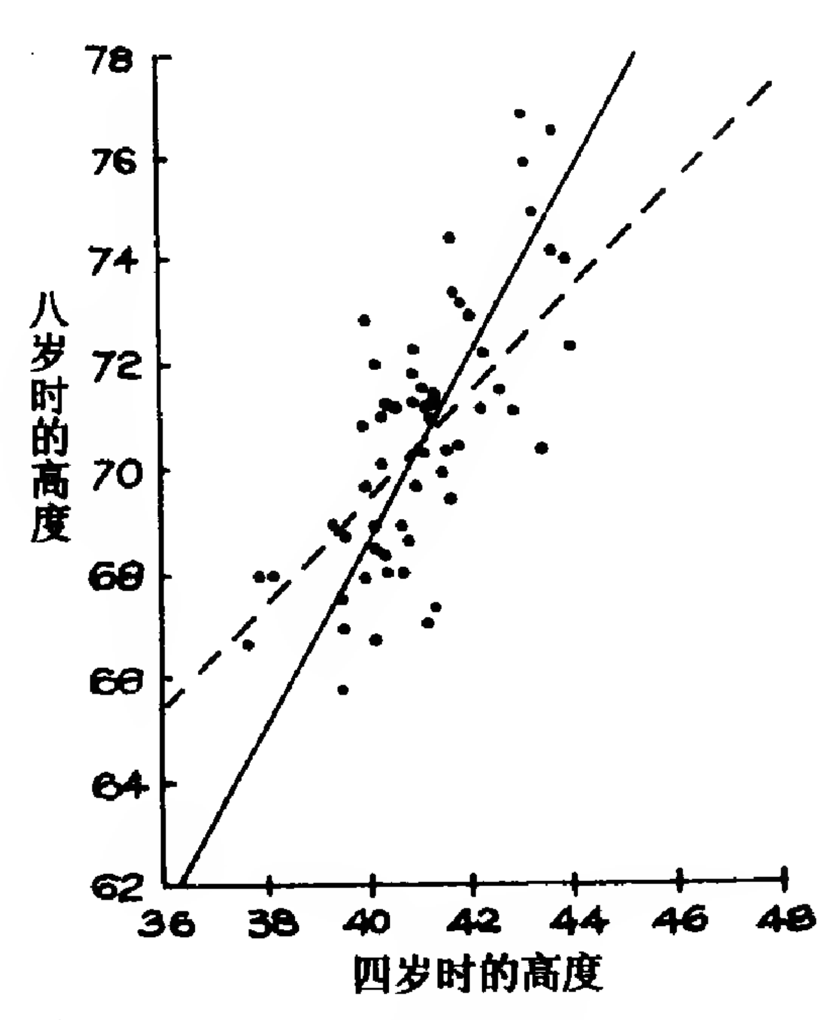
   (b)18岁时身高的SD约为

   0.5英寸　　1.0英寸　　2.5英寸

(c)相关系数约为

0.50　　0.80　　0.95

(d)哪一条是 SD 线——是实线还是虚线？对所选答案进行解释。



12. 某初级统计课程（在加利福尼亚大学柏克利分校）的十五个学生被要求数出类似如下页所示图中的点的数量（图中有 85 个点）。作数出点数的散点图，对每个学生散点面上有一点表示，它表明第一次数和第二次数的数量。需完整地标明两个坐标轴。可对坐标轴的尺度进行选择以便能看出点的型式。利用所作的散点图回答下列问题：

(a)所有学生都是独立地数吗？

(b)正确还是错误：第一次数得较多的学生第二次数也较多。

两次计数

第一次数	第二次数
91	85
81	83
86	85
83	84
85	85
85	84
85	89
84	83
91	82
91	82
91	82
85	85
85	85
87	85
90	85

6. 小结

1. 两个变量间的关系可直观地用散点图来表示。当散点图紧密地群集于一条直线的周围时，变量间存在强相关。

2. 一个散点图可用五个统计量来概括：

• 诸 x 值的平均数，诸 x 值的 SD，

• 诸 y 值的平均数，诸 y 值的 SD，

• 相关系数 r。

3. 正相关（点云斜向上）被相关系数前的正号指示，负相关（点云斜向下）被相关系数前的负号指示。

4. 在具有同样 SD 的一系列散点图中，当 r 接近于±1，点更紧密地群集于一条直线旁。

5. 相关系数的范围从－1（当所有点都在一条斜向下的直线上时）到＋1（当所有点都在一条斜向上的直线上时）之间。

6. 点群集在SD 线旁。该直线穿过平均数点，当 r 为正时，直线的斜率为

(y 的 SD)/(x 的 SD)

当 r 为负时，斜率为

−(y 的 SD)/(x 的 SD)

7. 为计算相关系数，将每个变量都转换为标准单位，然后取平均乘积。

9 再谈相关

“真的”，公爵夫人说：“火烈鸟和芥末都很刺鼻。那意思是说——‘物以类聚。’”

“但芥末并不是鸟”，Alice 说。

“是的，象往常那样，”公爵夫人说：“你具有多么清晰的表达方式！”

——《Alice 漫游奇境记》

1. 相关系数的特点

首先，相关系数纯粹是一个数。计算 r 的第一步是标准单位的转换，因此，变量的原始单位——如身高数据的英寸或温度数据的度——都被消去了。同样，如果把一个变量的所有值都乘以同一个正数，或者把某一变量的所有值都增加同一个数，r 都不会受到影响。（用统计学家的话说，r 不受尺度变化的影响。）

例如，如果把 x 的每一个值都乘以 3，则其平均数将扩大 3 倍，所有离平均数的偏差也将扩大 3 倍，因而 SD 亦扩大 3 倍。在标准单位转换中，消去了该公因子，所以 r 保持不变。又如，假定将

x 的每个值都增加 7，则 x 的平均数也将增加 7。但是，离平均数的偏差并不发生变化，r 也不会变化。

计算指出了有关散点图的一些性质。图 1 给出了纽约和波士顿的日最高气温间的相关。对 1985 年 6 月的每一天图中有一个点与之对应。纽约那天的温度用横轴表示，波士顿的温度用纵轴表示。左图是用华氏温度表示的，r＝0.69；右图是用摄氏温度表示的，r 保持不变[①]。

相关系数的另一个特点是：x 与 y 之间的相关系数等于 y 与 x 之间的相关系数。记住，r 是经过转换到标准单位后的乘积之平均数，而乘积的值并不受因子顺序的影响（a×b＝b×a）。再一次算术计算指出了散点图的某些性质。例如，图 2 中的左图给出的是 1985 年 6 月纽约市的日气温的散点图，每日的最低气温标在横轴；最高气温标绘在纵轴。最低与最高气温间的相关系数是 0.67。右图表示完全相同的数据；这一次最低气温标绘在纵轴而不是横轴。但 r 却是一样的，为 0.67。（令人吃惊的是相关系数仅为 0.67，不过，天气本身就是充满了意外的。）

> 相关系数是一个没有单位的纯数，它不受下列因素影响：
> - 互换两个变量；
> - 某一变量的所有值都增加同一数值；
> - 某一变量的所有值都乘以同一正数。

图 1 纽约和波士顿 1985 年 6 月的每日最高气温。左图是用华氏温度绘制的；右图用摄氏温度绘制。这对 r 毫无影响。

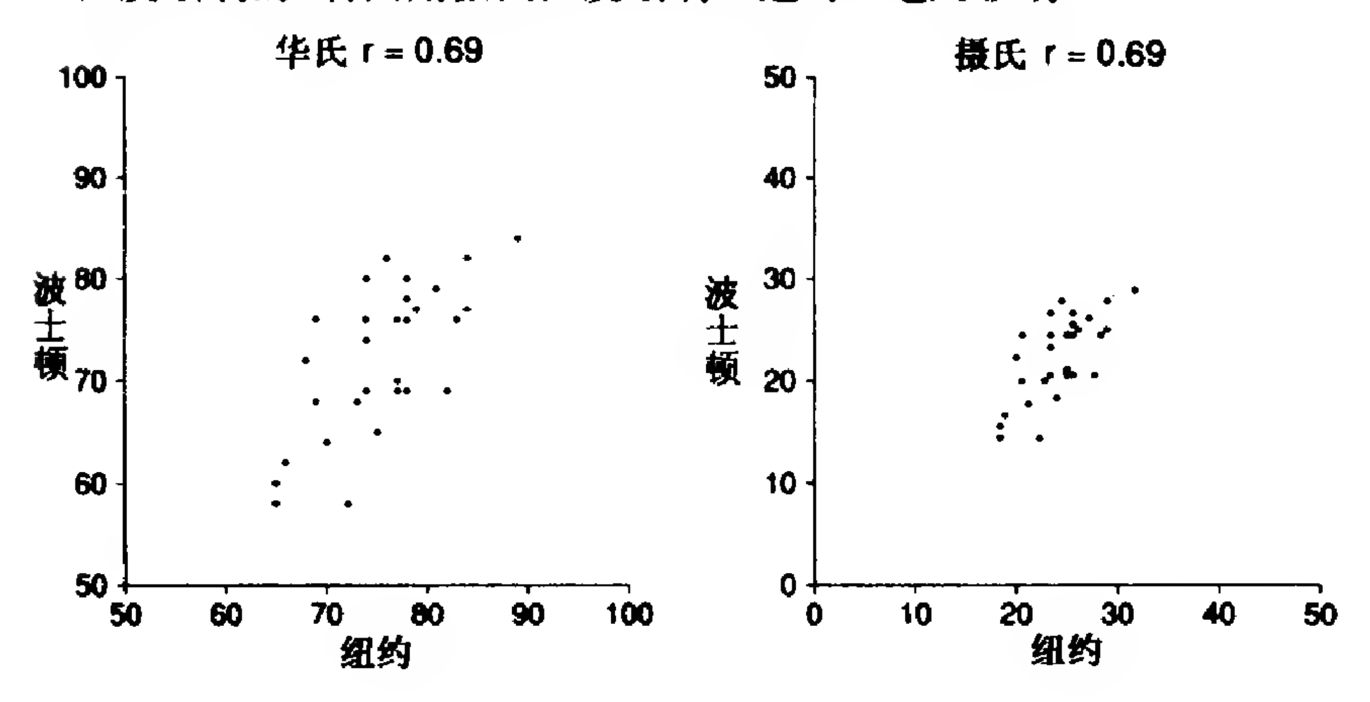
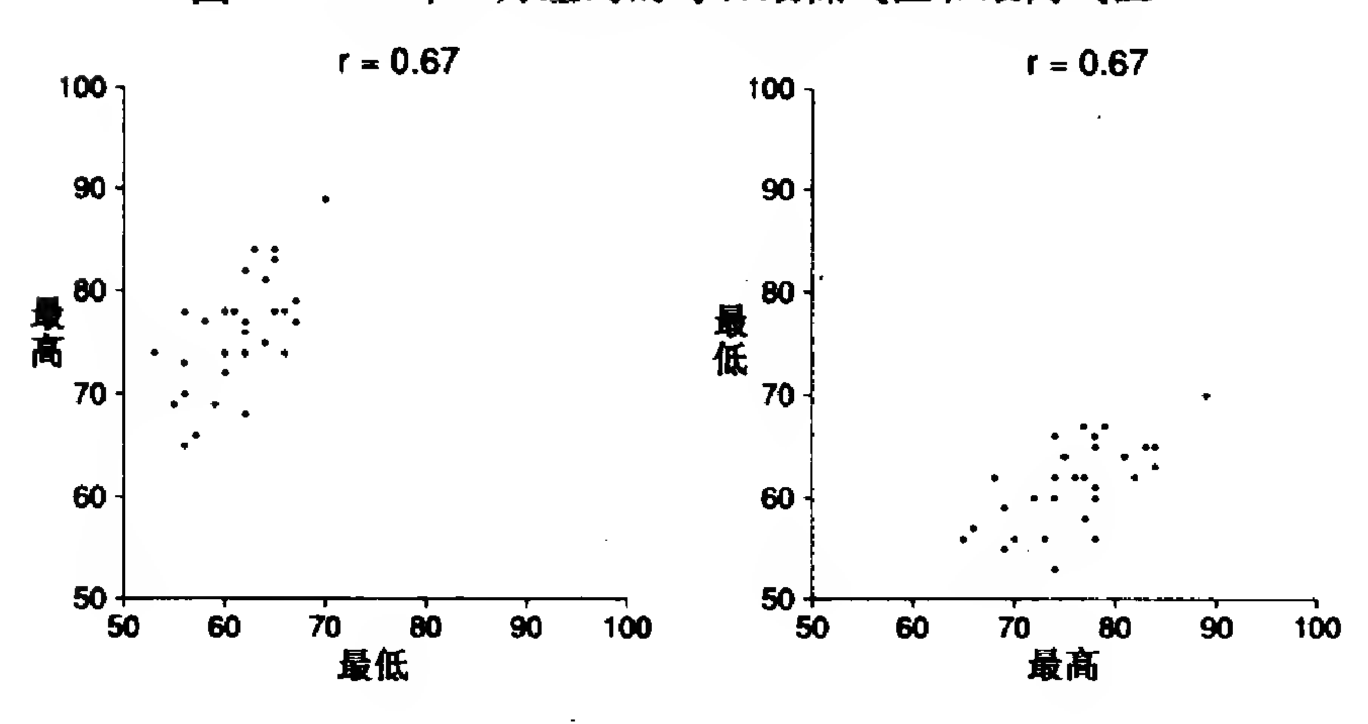


图 2 1985 年 6 月纽约的每日最低气温和最高气温。



习题 A

1. 在图 2 的左边部分，所有点都在 45 度线的上方，为什么？
2. 气象员计算了 1985 年 6 月华盛顿和纽约的日最高气温间的相关系数；然后又计算了华盛顿和波士顿间的相关系数。一个是 0.31 另一个是0.50。它们分别是哪两个城市间的相关系数？为什么？（在本题和下题中，“华盛顿”指城市，不是州。）

3. 两个气象员计算了华盛顿和纽约的日最高气温间的相关系数。一个计算的是 1985 年 6 月间的，另一个计算的是 1985 年全年的。谁得的相关系数较大？

4. 下面是一个小型的数据集，r≈0.76。若将两列交换，是否改变 r 的值？请解释或作计算。

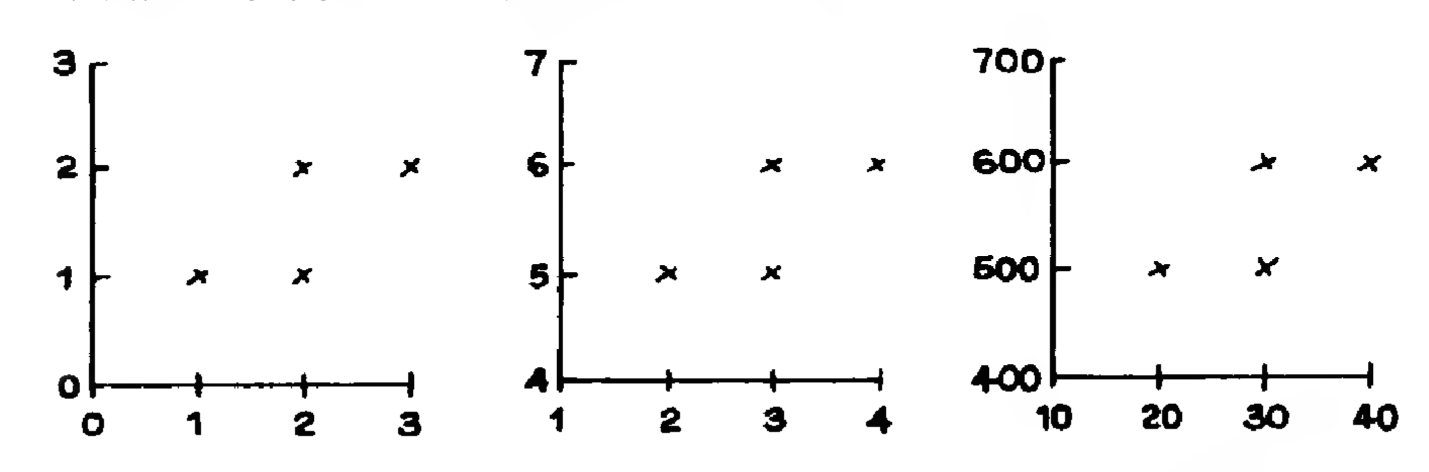
x	y
1	2
2	3
3	1
4	5
5	6

5. 如题 4，但将 y 的每一个值都加 3。

6. 如题 4，但将 x 的每一个值乘以 2。

7. 如题 4，但交换 y 的最后两个值(5 和 6)。

8. 下面是三个散点图，它们有相同的相关系数吗？试一下不作计算而回答。



9. 某人交给你如下所示的散点图，但忘了标明坐标轴。你仍能计算 r 的值吗？若能，该值为多少？或你需要标记吗？(见下图)

10. 两个不同的调研人员正进行一项关于生长的研究。第一位以英寸为单位测量了 100 个儿童的身高。第二位更喜欢公制，他把结果改为厘米(乘以转换系数 2.54 厘米/英寸)。

(a)如果转换不出差错，这两组测量值间的相关系数为多少？

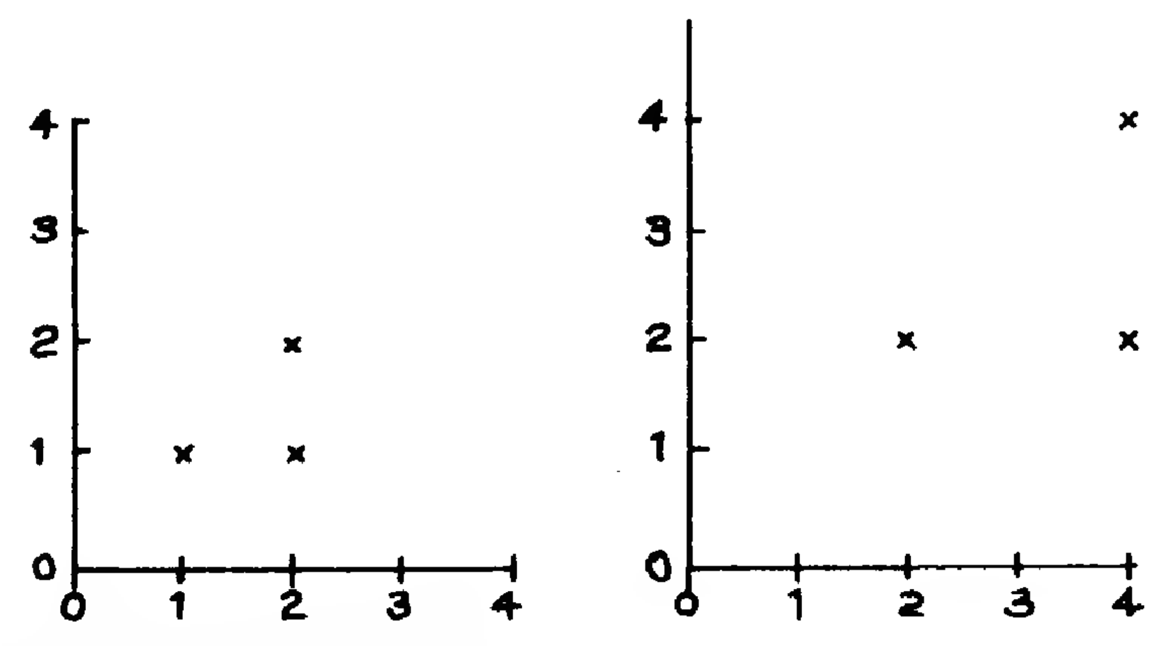
(b)如果转换中计算有误，r 将发生什么？

(c)如果第二位采用公制设备对上述儿童的身高再次进行测量，r 将发生什么？

11. 如下是五组数据。在(i)中，相关系数是 0.8571，在(ii)中为 0.7857。求出其余各组数据的相关系数。

(I)		(ii)		(iii)		(iv)		(v)		(vi)	
x	y	x	y	x	y	x	y	x	y	x	y
1	2	1	2	2	1	2	2	1	4	0	6
2	3	2	3	3	2	3	3	2	6	1	9
3	1	3	1	1	3	4	1	3	2	2	3
4	4	4	4	4	4	5	4	4	8	3	12
5	6	5	6	6	5	6	6	5	12	4	18
6	5	6	7	7	6	7	5	6	10	5	21
7	7	7	5	5	7	8	7	7	14	6	15

12. 下面是两个散点图。哪一个的相关系数较大？或二者的相关系数相同？请简短地说明。



13. 假设 x 与 y 间的相关系数为 0.73：

(a)散点图是斜向上还是斜向下？

(b)如果把所有 y 值都乘以－1，新的散点图是斜向上还是斜向下？

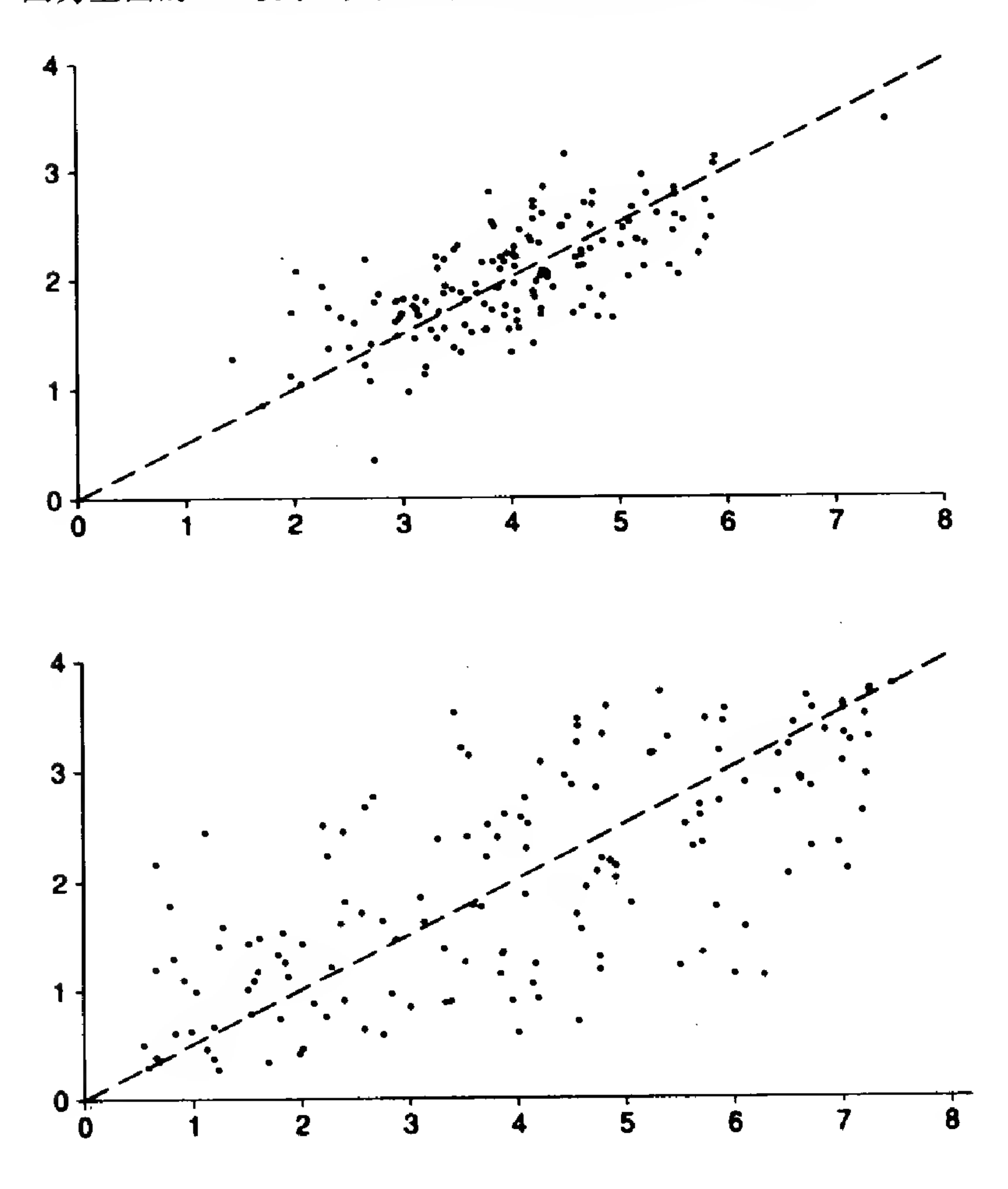
(c)如果把所有 y 值都乘以－1，相关系数将发生什么？

这些习题的答案在第 686—687 页上。

2. 改变 SD

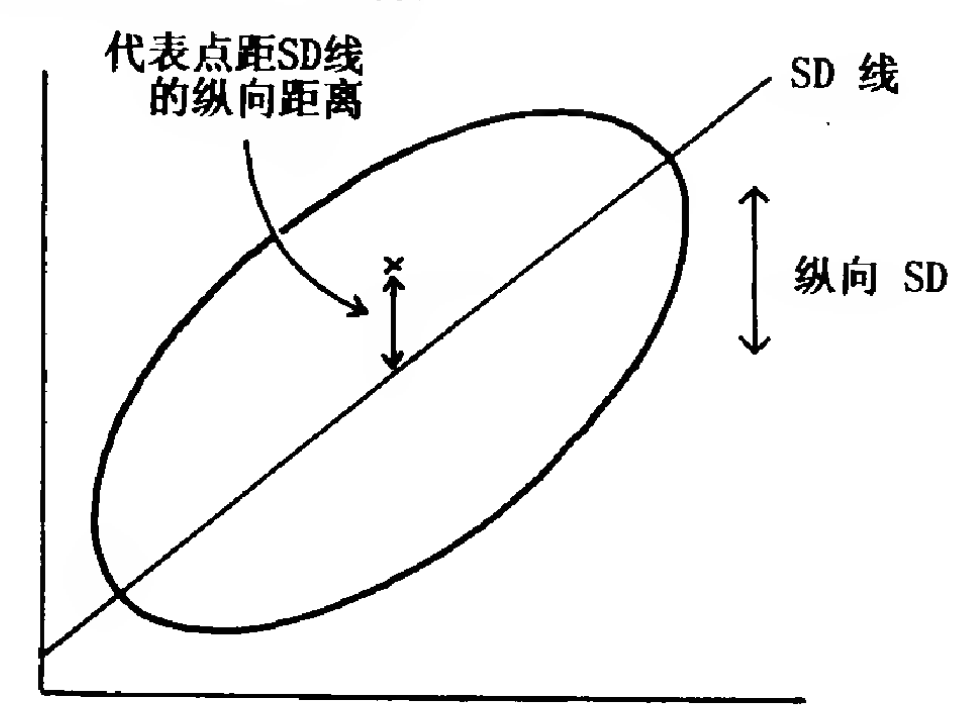
散点图的形状依赖于 SD。例如，考察图 3，在两个图中，r 都是 0.70。然而，上面一个图看起来围绕在 SD 线旁更密集些，这是因为它的 SD 较小一些。r 的计算公式涉及将变量转换为标准单位：离平均数的偏差除以 SD。结果是 r 不是按绝对值度量群集程度，而是按相对值——相对于 SD。

图 3　改变 SD 的效果。这两个散点图都有同样的相关系数 0.70。因为上图的 SD 较小，故看起来更密集地群集于 SD 线旁。



如果 r 接近于 1，那么一个代表点在 SD 线之上或之下的仅为纵向 SD 的一小部分的地方。如果 r 接近于 0，那么一个代表点位于 SD 线之上或之下一定的量，该量在大小上大致与纵向 SD 是可比较的（图 4）。对横向 SD 亦如此。

图 4 相关系数。当 r 接近于 1 时，代表点在 SD 线之上或之下的距离变成纵向 SD 的一小部分



为了用图形解释相关系数，可在你的心目中绘制一散点图，其纵向 SD 与第 142 页的图 6 相同，横向 SD 亦相同。如果你的散点图的 r 为 0.40，它的点与图 6 中 r 为 0.40 的右上图的点沿分角线群集程度大体一样。如果 r 为 0.90，则与图 6 中左下图大致相同。一般说来，你的散点图将匹配一个有同样 r 值的散点图。

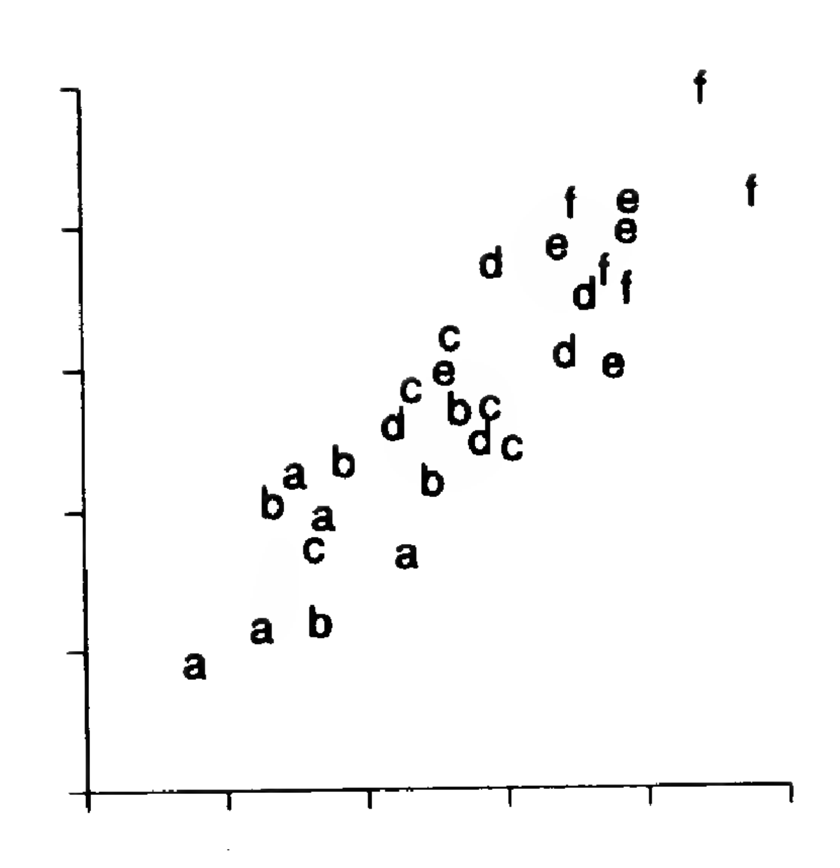
习题 B

1. 下图中，6 个散点图绘制在同一对坐标轴；第一个的点标以"a"，第二个的点标以"b"，余类推。单独考虑这 6 个散点图中的每一个，其相关系数是否大约为

   0.0　　0.6　　0.9

2. 续习题 1。把 6 个散点图的点合在一起，对这个合并起来的图，其相关系数是否大约为 0.0，0.6，或 0.9。
3. 健康和营养调查研究（HANES）也包括儿童。对 6 至 11 岁的每个年龄，身高和体重的相关系数约为 0.67。把所有儿童合在一起考虑，身高与体重的相关系数将恰好约为 0.67、比 0.67 稍大、还是比 0.67 略小？作一个选择并解释。

这些习题的答案在第 687 页上。



技术性注：相关系数与在SD线之上或之下的典型距离之间的联系可用如下数学式子表示：到SD线的r.m.s.纵向距离等于

$$\sqrt{2(1-|r|)} \times \text{纵向 SD}$$

例如，取相关系数为0.95，则

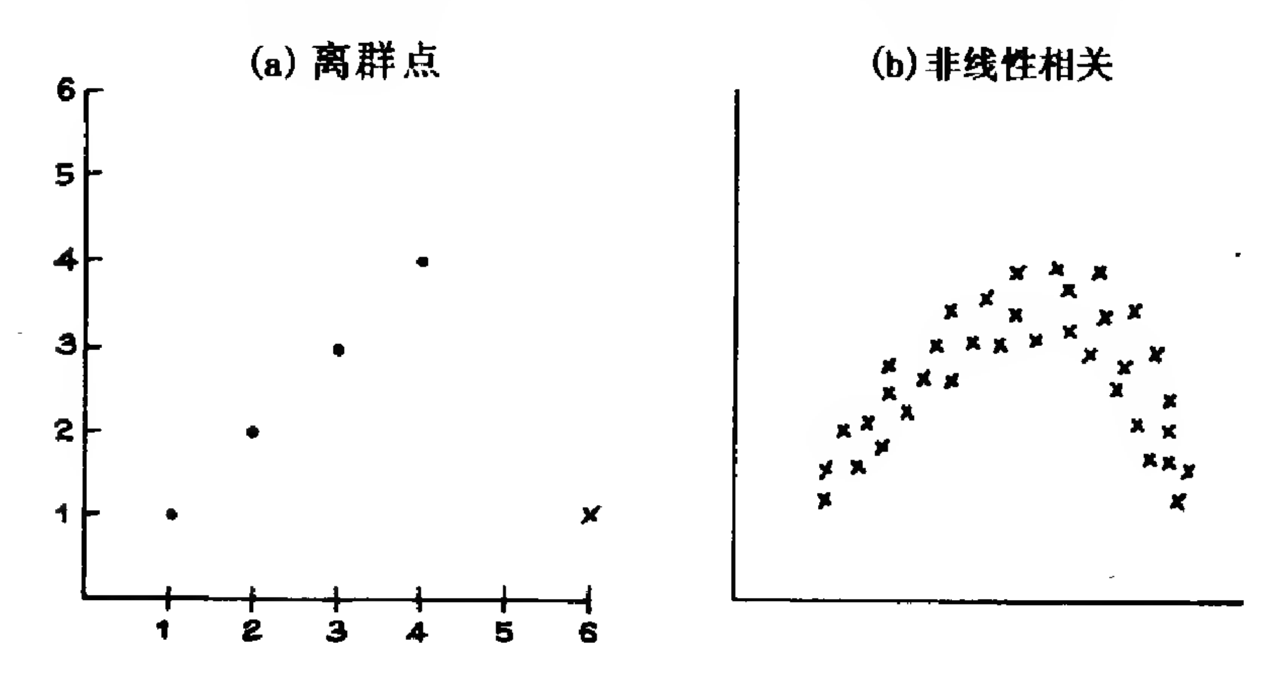
$$2\sqrt{(1-|r|)} = \sqrt{0.1} \approx 0.3$$

因此，在SD线周围的散布大约为纵向SD的30%。这就是为什么r=0.95的散点图显示了围绕SD线的相当量的散布的原因(P.142中的图6)。对水平方向也有同样的计算公式。

3. 一些例外情况

有时候用相关系数来描述散点图是一种很糟糕的办法。离群点和非线性就是两个成问题的情况。图5a中的点表示相关系数为1的完全相关，用叉表示的离群点使相关系数几乎减至0。图5a不应用相关系数来概括。(有人把离群点排除而不予考虑；然而，在任何一个散点图中，将总会有些点或多或少地游离于点云主体之外；仅当有理由时这些点被拒绝。)

图 5　离群点和非线性相关的存在可能使相关系数得出错误的结果。

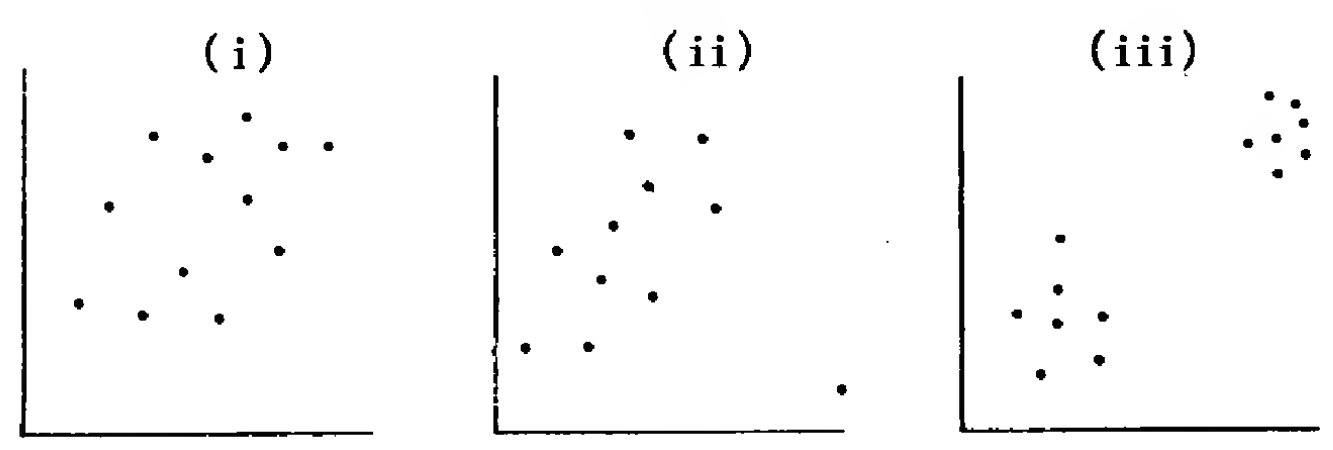


在图 5b 中，尽管点显示了强相关，相关系数仍接近于 0。理由是图象看来根本不象直线：当 x 增加时，y 先升后降。这种型式在考虑成人体重与年龄间的相关时即可见到（P66 图 3），再一次这种数据不能用相关系数 r 来概括——这种模式不存在。

> 只要可能，应通过观察散点图检查离群点和非线性相关：r 测量的是线性相关，不是一般意义的相关。

习题 C

1. 下面 3 个散点图中，哪一个应用 r 概括？



2. 一个有 15 个学生的班中恰好有 5 个篮球运动员。正确还是错误并解释

之：该班学生身高与体重间的关系能用 r 概括。

3. 对两个变量间的关系，散点图和相关系数哪一个给你的信息更多？请简短地解释。

这些习题的答案在第 687 页上。

4. 生态相关

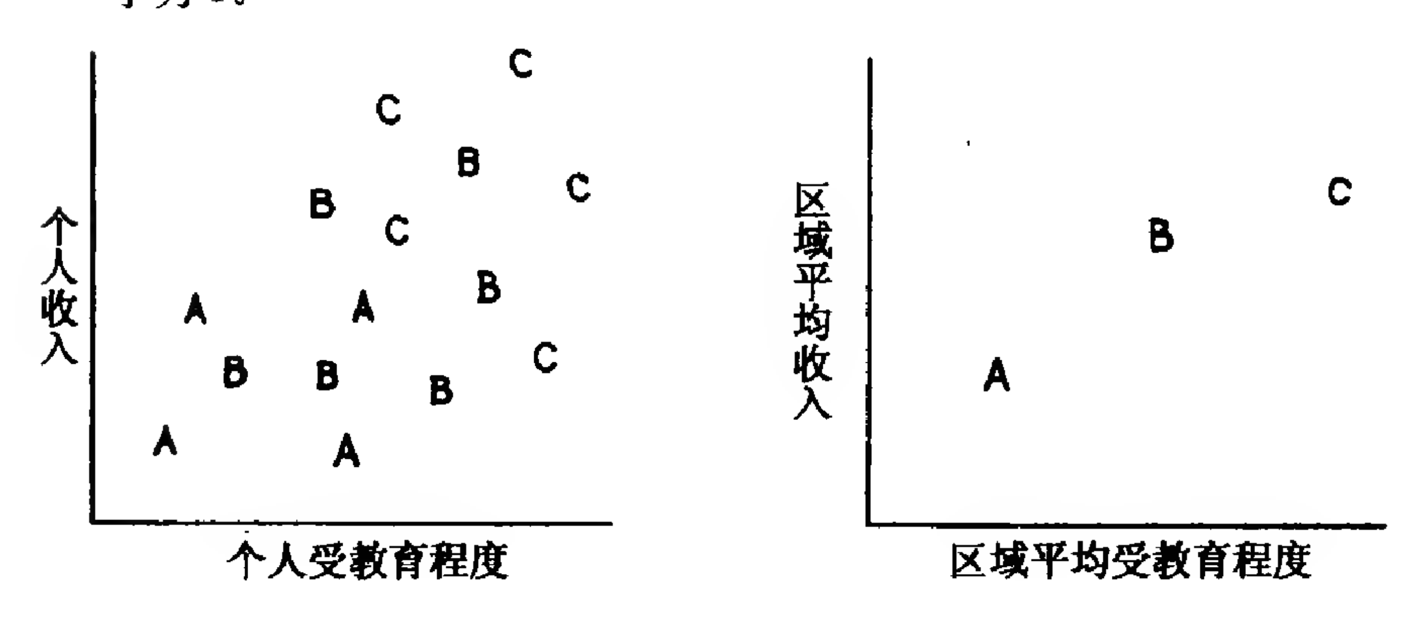
1955 年，Doll 发表了一篇划时代的文章，论述了吸烟与肺癌之间的关系[②]。其论据之一是在十一个国家中吸烟率（按人计算）与肺癌死亡率间关系的散点图。这十一对比率间的相关系数是 0.7，这被取作显示吸烟与肺癌之间关系的密切程度。然而，吸烟并得肺癌的不是国家，而是人。为了对人测量这种关系的密切程度，必须用关于个人吸烟和得肺癌的数据。这样的研究是有效的，并倾向于支持 Doll 的观点。

但是，基于比率或平均数的相关系数常使人误入歧途。这里有一个例子[③]。从 1988 年现场人口调查的数据，我们计算出美国 25—64 岁男子的收入与受教育程度间的相关系数：$r \approx 0.4$。普查局把美国分为 9 个地理区域。（例如，山区包括蒙大拿 Montana、爱德华 Idaho、怀俄明 Wyoming、科罗拉多 Colorado、新墨西哥 New Mexico、亚利桑那 Arizona、犹他 Utah 和内华达 Nevada）对这九个区域的任何一个，我们计算出居住在该区域的男子的平均收入和平均受教育程度。最后，我们计算出这九对平均数间的相关系数，其值为 0.7。如果你用区域的相关系数来估计个人的相关系数，你将大错特错。理由是在每一个区域内，围绕平均数有大量散布，用平均数取代一个区域则把这些散布都排除在外，给人以紧密群集的错误印象。图 6 对三个区域用图示说明了这一点。

> **生态**相关是基于比率或平均数的。它们通常用于政治科学和社会学领域。它们倾向于夸大相关程度，因此要小心一些。

图 6　基于比率或平均数的相关系数通常过大

左图表示了标为 A、B、C 的三个地理区域中个人的收入与受教育程度。每个人都用他所居住的区域的字母来标记。相关是中等程度的。右图表示的是每个区域的平均数。平均数间的相关系数几乎为 1。



习题 D

1. 下表摘自 Doll 的研究。表中数据为 1930 年各个国家人均烟消费量和 1950 年男子死于肺癌的比率。(1930 年时几乎没有任何妇女吸烟;吸烟的效果也需要一个较长的时间才能显示出来。)

国家	烟消费量	每百万人中的死亡数
澳大利亚	480	180
加拿大	500	150
丹麦	380	170
芬兰	1 100	350
大不列颠	1 100	460
冰岛	230	60
荷兰	490	240
挪威	250	90
瑞典	300	110
瑞士	510	250
美国	1 300	200

(a)绘出上述数据的散点图。

(b)正确还是错误:总的说来,上述国家中 1930 年的烟消费量较高的国家,

在 1950 年的肺癌死亡率也较高。或许从表中数据能得出这样的结论吗?

(c)正确还是错误:吸烟多的人的肺癌死亡率较高。或许从表中数据能得出这样的结论吗?

2. 某社会科学家正在对十九世纪意大利自杀和有文化之间的关系进行研究[4]。他有各个省的数据,给出了各省有文化的人的百分比与自杀率。相关系数是 0.6。这对有文化与自杀之间的相关程度给出了合理的估计吗?

这些习题答案在第 687 页上。

5. 相关不是因果关系

对于在校儿童,鞋的大小强烈地与阅读技能相关联。然而,学会新词并不会使脚变大一些。而是涉及到第三个因素——年龄,当儿童长大一些,他们学会了阅读得更好且由于长大而穿不下原来的鞋。(按第 2 章的统计术语,年龄是一个混杂因素。)在本例中,混杂因素很容易被看出;在大多数情况下,事情就没有这么简单了。相关系数的算术运算,并不能保证你不受第三因素影响[5]。

> 相关系数度量的是相互关系,但相互关系并不等于因果关系。

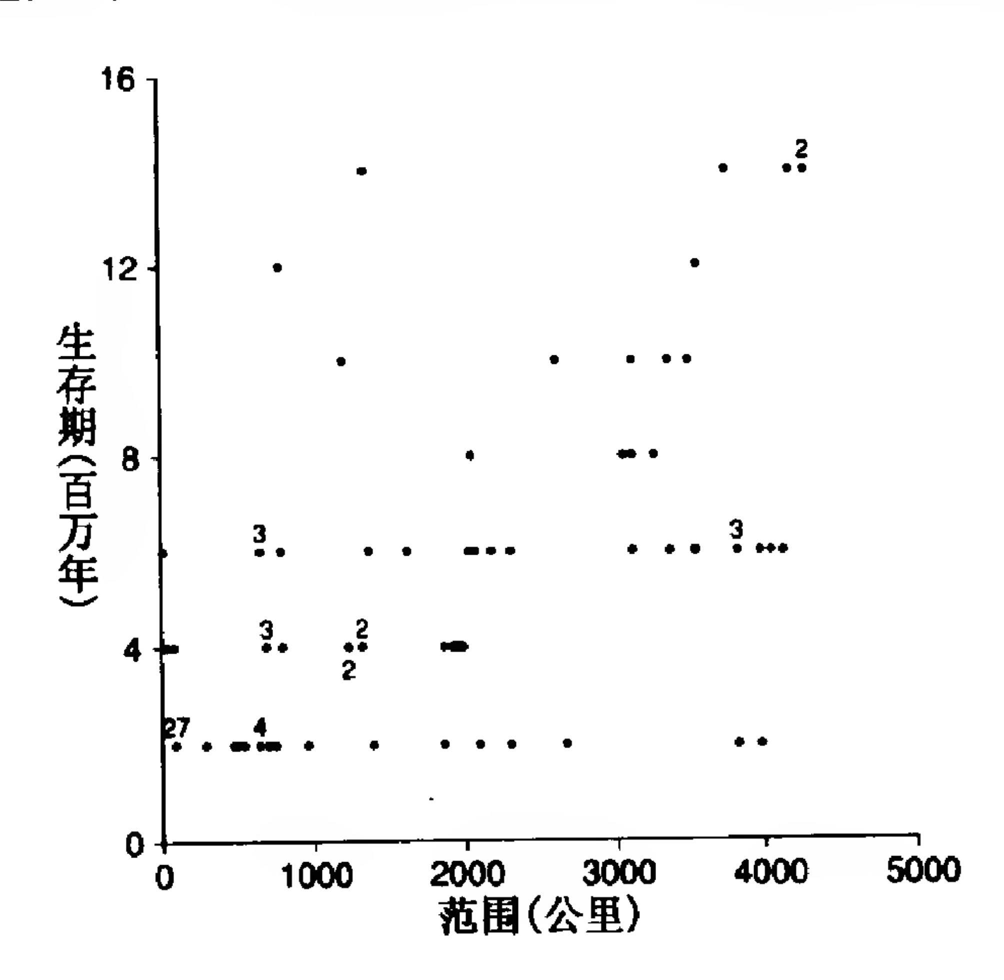
例 1. 受教育程度与失业。大萧条期间(1929—1933),受过良好教育的人趋于失业时间较短。教育能使你免于失业吗?

讨论:也许如此,但数据是观察性的。事实上,年龄是混杂变量。由于教育水平一直在提高(现在亦如此),越年轻受的教育越好。在作雇人选择时,雇主总是倾向于要年轻求职者。控制年龄就会使教育对失业的影响大大减弱[6]。

例 2. 物种生存范围与生存期间。自然选择对物种级别有作用吗?这是古生物学家感兴趣的问题。David Jablonski 提出地理范围是物种的遗传特征:生存范围较宽的物种生存期较长,这是因为如果灾难在某地降临,该物种在另一地仍然存在。

证据之一是散点图(图 7),图中有九十九种腹足纲软体动物(蛞蝓、蜗牛,等等)。物种的生存期——其存在时期(以百万年为单位)——标于纵轴,其生存范围(以公里为单位)标于横轴。这两个变量都取自化石记录。它们之间存在较强的正相关:r 约为 0.64。(点云看起来杂乱无章,而这只是因为散落在右下方及左上方的少数点所致。)宽的地理范围有助于物种的生存吗?

图 7 99 种腹足纲软体动物按百万年计的生存期与以公里计的地理范围。同一个点可能表示几个物种,这时物种数目标在点旁。

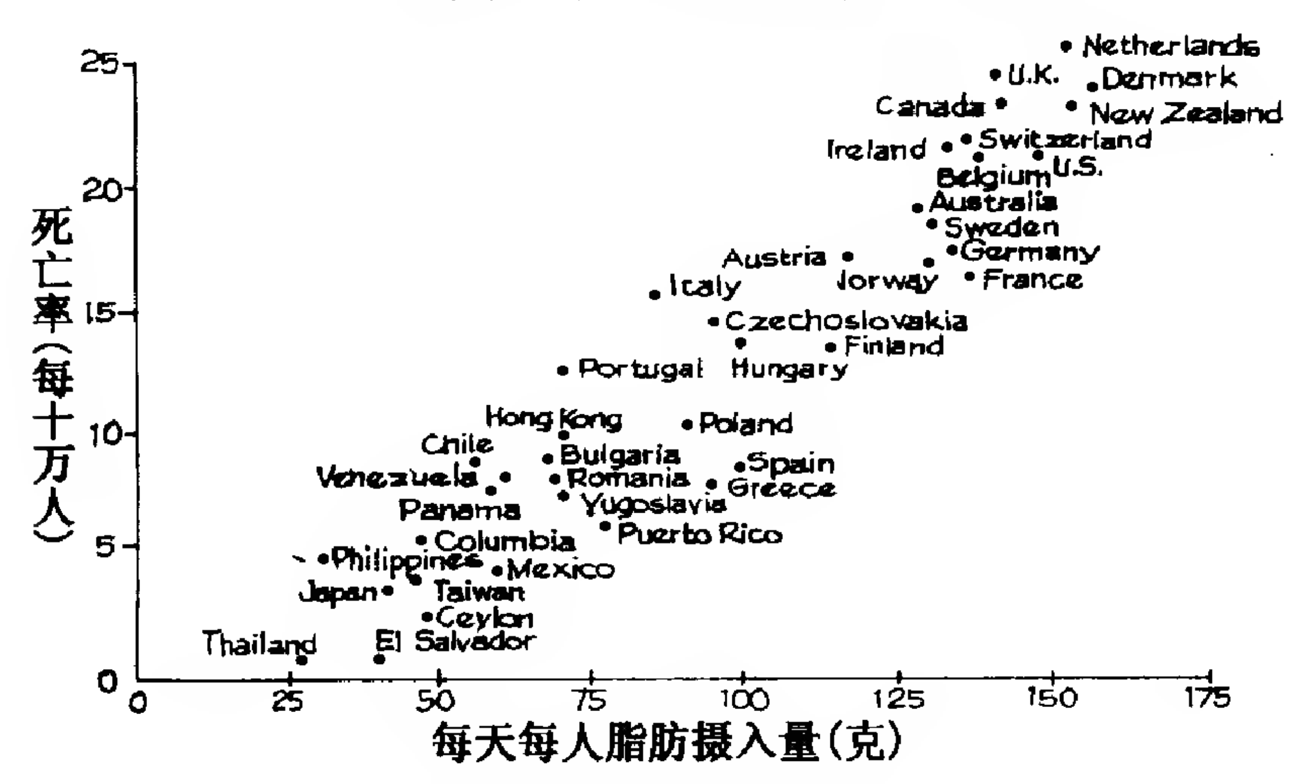


讨论:物种宽的生存范围可能使其生存期长;或者,长的生存期可能引起宽的生存范围;或者,可能还有其他因素。Jablonski 注意到第一种可能性。第二种不太可能,因为其他证据表明物种出现后将很快布满其生存范围。那么第三种解释呢? Michael Russell 和 David Lindberg 指出,生存地理范围宽的物种其化石记录保存下来的可能性更大,这给人以其生存期长的印象。如果如此,图 7 只是一个人为的统计现象⑦,相关并不等同于因果。

例 3. 饮食中的脂肪与癌症。在人们摄入脂肪较多的国家——如美国——乳腺癌与结肠癌的发病率很高。见图 8。这种相关性常被用来说明饮食中的脂肪引起癌症。这种论据站得住脚吗?



图 8 一组样本国家的癌症发病率与脂肪摄入量。



来源:K. Carroll,"Experimentnl evidence of dietary factor and hormonc—dependent Cancers." Cancer Research Vol. 35(1975) P. 3379.

讨论:如果摄入脂肪导致癌症,则在其他条件相同的情况下图中的点呈上升趋势,因此此图是该理论的某种证据。但这种证据是很弱的,因为其他条件并不相同。例如,饮食中含很多脂肪的国家亦有很多的糖。结肠癌发病率与糖的消费量之间的散点图看上去正好与图 8 相象,却没有人认为糖导致结肠癌。事实上,脂肪和糖都是比较昂贵的。在富国,人们吃得起脂肪和糖,吃淀粉含量高的粮食制品少。这些国家饮食结构的某些方面,或生活方式的其他因素,可能会导致某些癌症——也可能防止其他癌症。到目前为止,流行病学家只有把握证实这些因素中的少数几个,脂肪不在其中[8]。

习题 E

1. 图 7 中的散点图显示出一些条状,为什么?

2. 图 8 的相关是生态相关吗？这一点怎样能与论点相联？

3. 美国 18—74 岁男子身高与体重之间的相关系数约为 0.40。正确还是错误，并给予解释：

(a)较高的男子趋于较重。

(b)美国 18—74 岁男子体重与身高之间的相关系数为 0.40。

(c)体重较重的男子趋于较高。

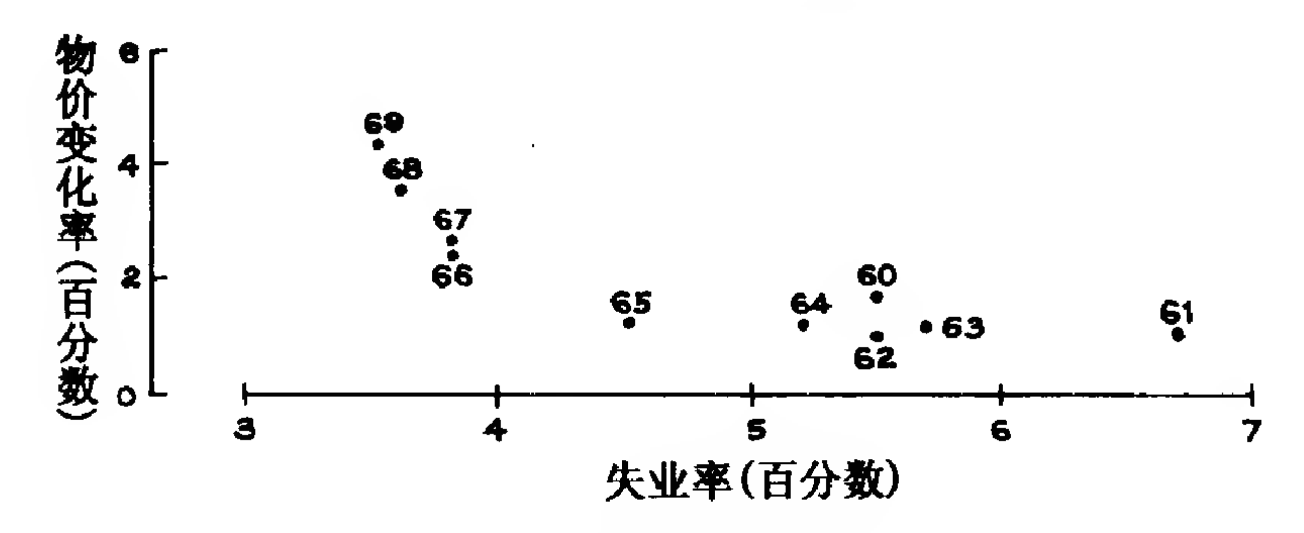
(d)如果多吃一些从而增加 10 磅体重，你的身材会长高。

4. 研究发现花在看电视上的小时数与阅读测验得分之间是负相关[9]。看电视真会使阅读能力降低吗？请简短地说明。

5. 许多研究发现吸烟与心脏病间存在相关。某研究发现喝咖啡与心脏病之间有相关关系[10]。你将得出喝咖啡会导致心脏病的结论吗？或者你能用其他方式解释喝咖啡与心脏病间的相关关系吗？

6. 许多经济学家相信失业与通货膨胀间存在权衡关系：低失业率将导致高通货膨胀率，高失业率将降低通货膨胀率。美国 1960－1969 十年中这两个变量间的关系如下图，对每一年都有一个点，失业率表示在 x 轴上，通货膨胀率表示在 y 轴上。这些点是下降的并接近一条称为菲利普曲线的平滑曲线。这是观察研究还是对照实验？如果绘出 70 年代或 50 年代的点，你期望它们会沿此曲线下降吗？


60 年代的 Phillips 曲线



来源：Economic Report of the President (1975)

这些习题的答案在第 687—688 页上。

6. 复习题

下列复习题可能包含前面各章的内容。

1. 研究单个变量问题时，你可利用称为________的图。研究两个变量间的关系时，你可利用称为________的图。
2. 正确还是错误，并简短地说明：
   (a)如果相关系数为－0.80，小于平均数的因变量值与小于平均数的自变量值相关。
   (b)如果 y 一般都小于 x，x 与 y 之间的相关系数将是负值。
3. 下面各种情形中，指出哪一个相关较高并简短地说明。(数据取自关于生长的纵向研究。)
   (a)4 岁时的身高和 18 岁时的身高，16 岁时的身高和 18 岁时的身高。
   (b)4 岁时的身高和 18 岁时的身高，4 岁时的体重和 18 岁时的体重。
   (c)4 岁时的身高和体重，18 岁时的身高和体重。
4. 一项研究中男大学生身高和体重间的相关系数约为 0.60，女大学生的情况也大致如此。如果把男生和女生合在一起考虑，身高与体重间的相关系数将为________。

   大约 0.60　　　　略小一些　　　　略大一些

   选择一个答案并简短地解释之。
5. 为了测定湖中水的清洁程度，将一有刻度线的玻璃片放入水中直至完全看不见刻度线；此时它与水表面的距离称为"Secchi 深度"。为了测量湖水被水藻污染的程度，科学家要确定水中叶绿素的总浓度。在某一湖中，从四月至九月每周四中午都测量 Secchi 深度和叶绿素的总浓度。这两个变量间是正相关还是负相关？请简短地说明。
6. 在下面的各数据集中都有一个数遗漏。如果可能，填空以使 r 等于 1。如果这不可能，请说明为什么。

(a)

x	y
1	1
2	3
2	3
4	—

(b)

x	y
1	1
2	3
3	4
4	—

7. 一计算机程序打印出了下面两个数据集的 r，这个程序运行是否正确？回答是或否，并简短地说明。

(i)

x	y
1	2
2	1
3	4
4	3
5	7
6	5
7	6

r＝0.8214

(ii)

x	y
1	5
2	4
3	7
4	6
5	10
6	8
7	9

r＝0.7619

8. 1910 年，Hiram Johnson 参加了加利福尼亚(California)州的州长竞选。每个县的数据都显示了该县土生土长美国人的百分比，及 Johnson 的支持率。一政治学家计算了这两个百分率之间的相关系数[11]，为 0.5。能否认为这准确地测量了"Johnson 获得了土生土长美国人的支持，和移民的反对?"的程度。回答是或否，并简短地说明。

9. 对美国 1988 年 25 岁及以上的男子，其年龄与受教育水平(完成正规教育的年数)间的关系可概括如下[12]：

   平均年龄≈47 岁，SD≈16 岁

   平均受教育水平≈12.5 年，SD≈3.5 年，r≈－0.27

正确还是错误并作解释：随着年龄的增长，受的教育会减少。如果这个结论不对，负相关的原因是什么？

10. 在 California 大学 Berkeley 分校，统计学 2 是分成助教主持的若干个小型讨论班的大课。作为一项研究的一部分，在学期的倒数第二课时，要求学生填写不记名问卷，对他们的助教(记名)的教学效果和课程进行评定，评价尺度如下：

1	2	3	4	5
差	可以	好	很好	优秀

计算出来下述统计量：

- 各小班学生对助教的平均评价。
- 各小班学生对课程的平均评价。
- 各小班学生的平均期末考试分数。

计算结果如下，各班用字母标记。绘出各对变量——共有 3 对——的散点图并判断其相关关系。

（这些相关关系是基于平均数的。因为问卷是不记名的，所以把各个学生的评价与考试分数联系起来是不可能的。学生的能力可能是一个混杂因子；然而，与以前的考试成绩相对比可消除分析的差异[13]。一个助教只教一个班。）

正确还是错误，并解释：

(a)平均说来，比较喜欢他们的助教的班的学生期末考试成绩较好。

(b)各班对助教的平均评价与对课程的平均评价间几乎没有关系。

(c)各班对课程的平均评价与平均期末考试成绩间几乎没有关系。

班别	对助教的平均评价	对课程的平均评价	期末考试平均成绩
A	3.3	3.5	70
B	2.9	3.2	64
C	4.1	3.1	47
D	3.3	3.3	63
E	2.7	2.8	69
F	3.4	3.5	69
G	2.8	3.6	69
H	2.1	2.8	63
I	3.7	2.8	53
J	3.2	3.3	65
K	2.4	3.3	

11. 1983 年，调研人员对在纽约的俄国流亡者进行了问卷调查[14]。

认为克格勃领导人“很有能力”的人同时也说他们很少参加俄国的政治抗议活动。这种相关在对年龄、受教育程度和职业进行调整后仍然存在。调研人员得出结论：

…认为克格勃能力很强的人不太可能卷入非正统行为……对克格勃能力的这种认识可能威慑不守规矩的人。

这个结论合理吗？回答是或否，并简短地说明。

12. 在对丹麦应征入伍者的研究中，T. W. Teasdale 及其助手们发现近视程度与智力间有正相关关系[15]。正确还是错误，并解释：

(a)一般说来，近视程度高的应征入伍者智力也高。

(b)一般说来，智力高的应征入伍者近视程度也高。

(c)这些数据说明近视导致高智力。

(d)这些数据说明高智力导致近视。

13. 在对 1988 年 SAT 数学分数的研究中，教育考试处计算了五十个州中每个州的平均分数，及该州最高年级中学生参加这个考试的百分比[16]。两个变量间的相关系数是－0.86。

(a)如何解释这个负相关现象?

(b)纽约州的平均分数只有 469；而怀俄明州的平均分数是 527。正确还是错误，并解释：这些数据表明，一般说来，怀俄明州的中学比纽约州的中学教学质量高。

7. 小结

1. 相关系数纯粹是一个数，没有单位。它不受下列情况影响：
   - 交换两个变量。
   - 对一个变量的所有值都增加同一个数量。
   - 把一个变量的所有值都乘以一个正数。

2. 相关系数度量点围绕直线群集的程度，但这只是相对于 SD 而言的。

3. 在有离群点或非线性相关的情况下，相关系数可能使人误入歧途。只要可能，应观察散点图以检查这类问题。

4. 生态相关，这是基于比率或平均数的相关，会夸大相关的程度。

5. 相关系数度量相关关系。但相关关系并不一定表示因果关系。它可能只说明两个变量同时都受某第三变量影响。

10
回归

你必须在某个地方绘出这条线。

1. 引言

回归方法描述一个变量如何地依赖于另一个变量。以身高和体重为例。我们有 988 个 18—24 岁男子的数据(数据取自健康和营养检查调查——HANES;见 64 页),他们的平均身高为 70 英寸,总平均体重为 162 磅。自然地,高的男子重一些。当身高增加一个单位时相应地体重增加多少呢?首先来看散点图(图 1),身高在横轴,体重在纵轴。概括统计量为[1]:

平均身高≈70 英寸,SD≈3 英寸

平均体重≈162 磅,SD≈30 磅,r≈0.47

选取纵轴和横轴的尺度使身高的一个 SD 与体重的一个 SD 在图中是同样长,这样 SD 线(虚线)在图中以 45 度上升。大量散点分布于此线旁;r 只有 0.47。

图 1 身高和体重的散点图。图中每一点代表健康与营养检查调查中 988 个年龄在 18－24 岁男子之一的身高和体重。纵向条形图中的点代表身高在平均身高之上约一个 SD 的所有男子，他们中体重也在平均体重之上一个 SD 的男子落在 SD 线上。这些男子的大多数都在 SD 线之下，因此他们的体重只是平均体重之上一个 SD 的一部分。以实线绘出的回归直线估计了每一身高的平均体重。

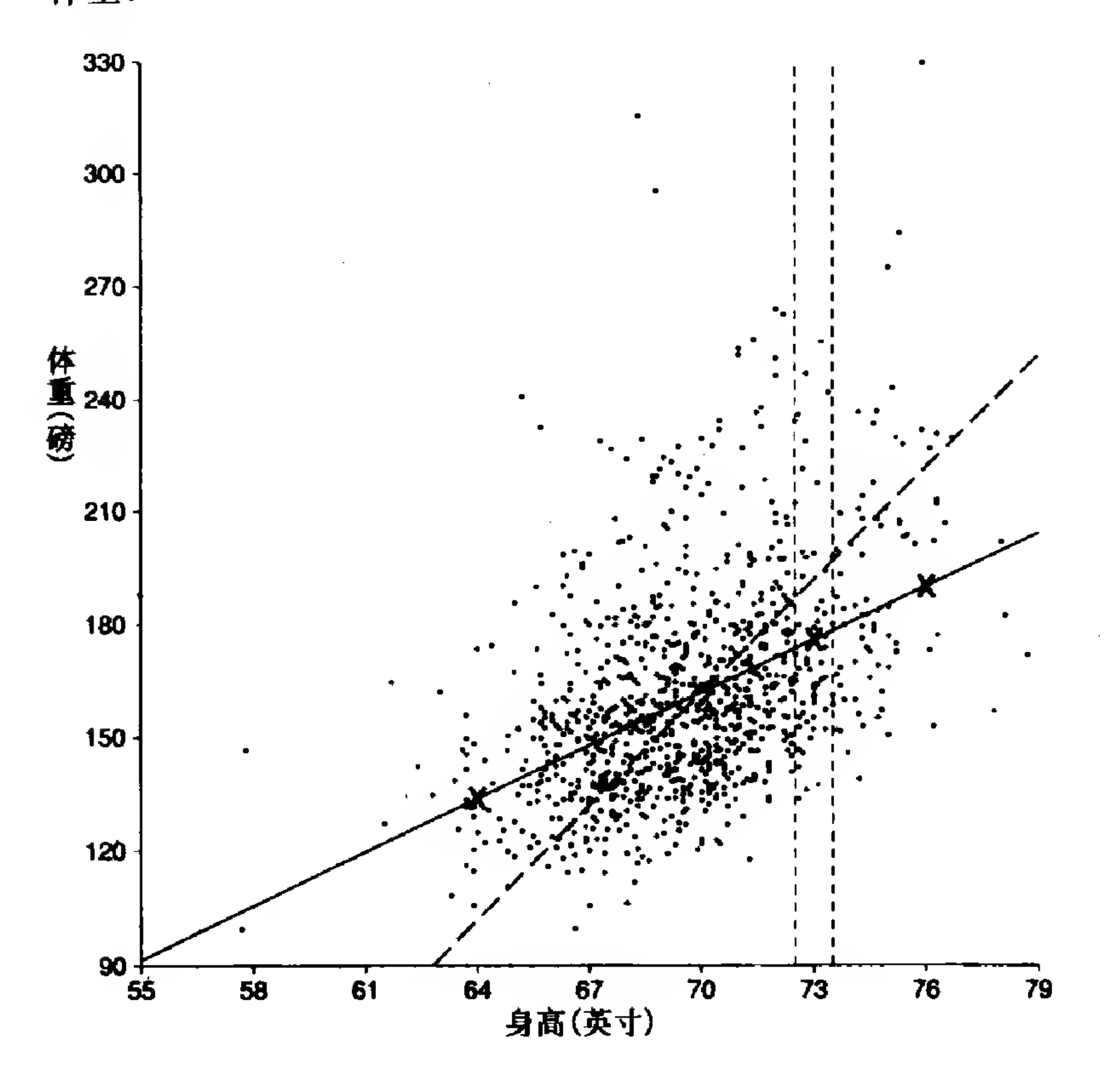


图 1 中的纵向条形图表示身高在平均身高之上一个 SD(近似到英寸)的男子，他们中体重也在平均体重之上一个 SD 的男子将沿 SD 线标绘。但是，这个纵向条形图中的大多数点都在 SD 线之下。换句话说，身高在平均身高之上一个 SD 的大多数男子的体重都比平均体重之上一个 SD 要小一些，因此这些男子的平均体重只

是总平均体重之上一个 SD 的一部分，哪一部分？这就要看相关系数 0.47 了。与身高增加一个 SD 相应，体重平均只增加 0.47 个 SD。

更准确些说，对于身高在平均身高之上一个 SD 的男子：

平均身高＋身高 SD＝70 英寸＋3 英寸＝73 英寸

他们的平均体重在总平均体重之上 0.47 个体重 SD。换算为磅，为

$$0.47 \times 30 \text{ 磅} \approx 14 \text{ 磅}$$

因此他们的平均体重约为

$$162 \text{ 磅} + 14 \text{ 磅} = 176 \text{ 磅}$$

他们的点(73 英寸，176 磅)在图 1 中用叉来表示。

对身高在平均身高之上 2 个 SD 的男子又如何呢？现在

平均身高＋2 个身高 SD＝70 英寸＋2×3 英寸＝76 英寸

这第二组男子的平均体重应在总平均体重之上 0.47×2＝0.94 个体重 SD，即 0.94×30 磅≈28 磅。因此，他们的平均体重约为 162 磅＋28 磅＝190 磅。点(76 英寸，190 磅)在图 1 中也用一个叉表示。

对身高在平均身高之下 2 SD 的男子又如何呢？他们的身高等于

平均身高－2 个身高 SD＝70 英寸－2×3 英寸＝64 英寸

他们的平均体重在总平均体重之下 0.47×2＝0.94 个体重 SD，即 0.94×30 磅≈28 磅。因此，这第三组的平均体重约为 162 磅－28 磅＝134 磅。图 1 中的第三个叉就是点(64 英寸，134 磅)。

所有(身高，平均体重估计值)这样的点都落在图 1 中的实线上，这就是回归线，它穿过平均数点，即平均身高的男子应也有平均体重。

回归线对散点图就象平均数对数据表一样。y 关于 x 的回归线估计了相应于每一个 x 值的 y 的平均数。

在回归线上，身高每增加一个SD，相应地体重只增加0.47个SD。具体地说，设想把这些男子按身高分组，其中有一个组具有平均身高，另一组是平均身高之上一个SD，如此等等。从一个组到另一个组，平均体重也将增加，不过只增加约0.47个SD。注意0.47是怎么得来的：它是身高与体重之间的相关系数。

这种用相关系数估计每个x值所对应的y的平均数的方法叫回归方法。这种方法可叙述如下：

> x每增加1个SD，平均而言，相应的y增加r个SD。

这里涉及两个不同的SD：x的SD，测定x的变动范围；y的SD，测定y的变动范围。很容易被如下的韵律扰昏了头：如果x增加1个SD，y亦如此。但这是错误的。一般说来，y只增加r个SD。请看图2。

为什么r是正确的因子呢？要证明这一点需要复杂的数学讨论。但三种情况容易直接看出。首先，假设r为零，即x与y之间不相关，因此x增加1个SD，相应的y平均增加0SD。其次，假定r为1，则所有点都落在SD线上，x每增加1个SD相应地y也增加1个SD。第三，假定r为−1，除了直线向下倾斜外，其他论述都与第二种情况相同。

对于中间r值的情况，我们来看图3，这是r=0.50的散点图。x和y的平均数都是4，SD都为1。SD线以45度上升，沿着此线，x每增加1个SD则y也增加1个SD。现在取5上方的纵向条中心点，那里x值约在平均数之上约1个SD。该纵条中点的y的平均数标以叉，它正好位于穿过y的平均数的水平线与倾斜的SD线之间的当中。换句话说，x每增加1个SD，平均来说y相应地增加半个SD。这再一次说明，r是正确的因子。

图 2　回归方法。当 x 增加 1 个 SD，y 的平均值只增加 r 个 SD。

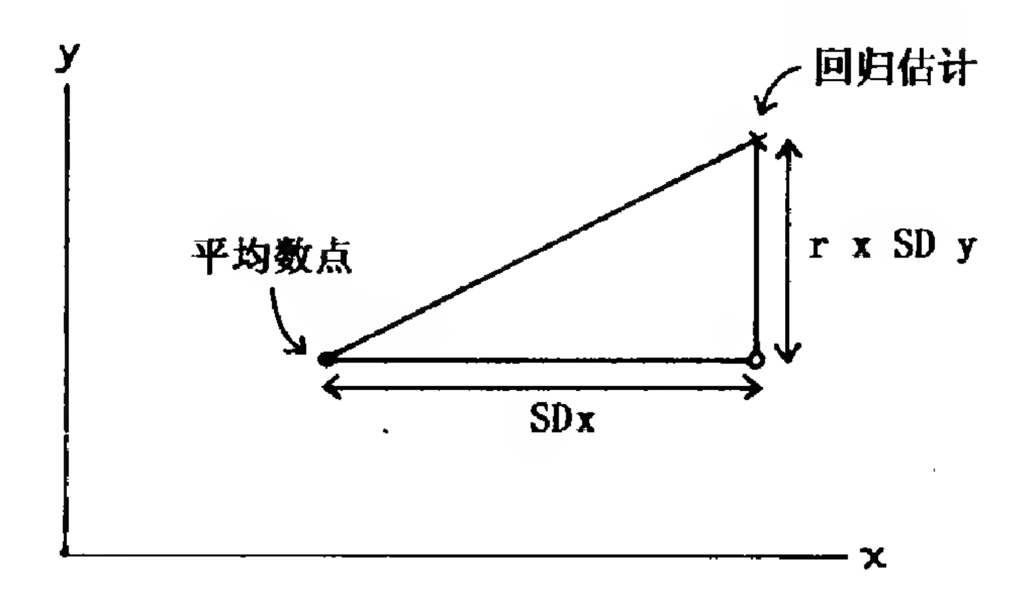
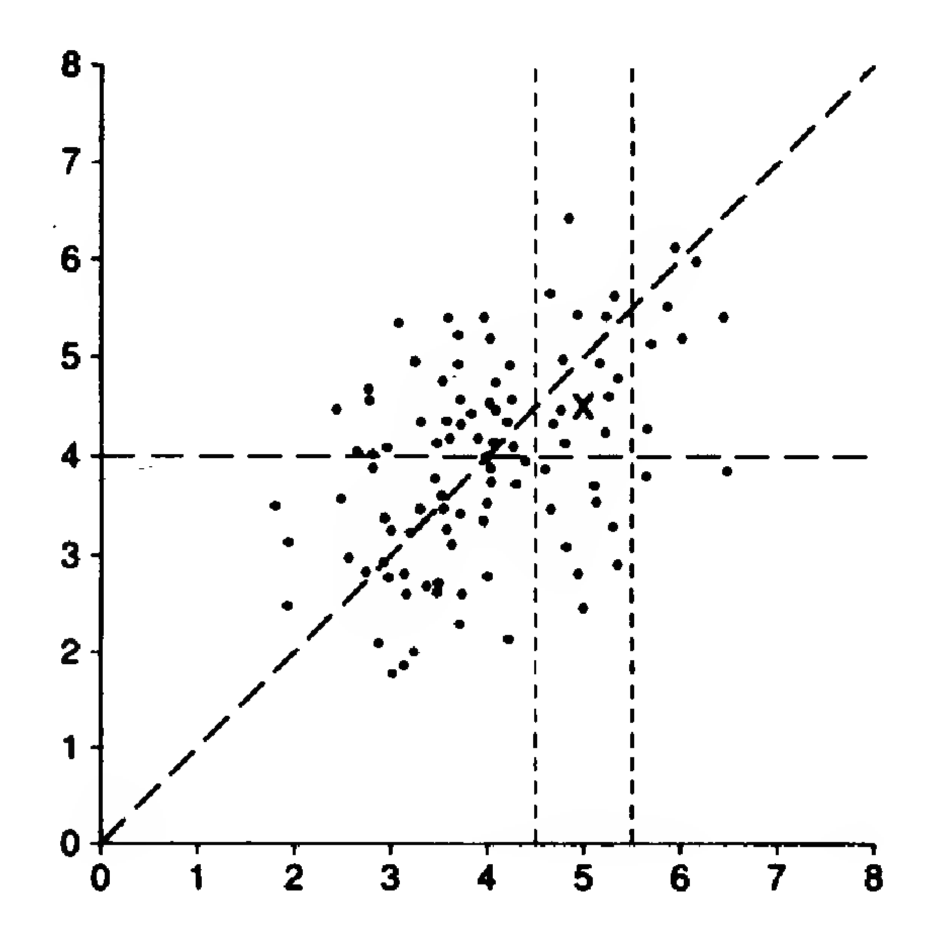


图 3　r＝0.50 的散点图。x 和 y 的平均值都是 4，SD 都是 1。5 上方的纵向条形图中的点其 x 值在平均数之上 1 个 SD。它们的 y 值的平均数用叉表示，它正好在穿过 y 的平均数的水平线与倾斜的 SD 线之间的当中。因此，x 每增加一个 SD，平均来说，y 相应地增加 r 个 SD。



习题 A

1. 对美国 1988 年 25－34 岁的男子，受教育程度（完成正规学校教育年数）与个人收入之间的关系可概括如下[2]：

   平均受教育水平≈13 年，SD≈3 年

   平均收入≈21 800 美元，SD≈15 000 美元，r≈0.34

   试估计小学毕业但未进入中学的男子（因此他们受教育年数为 8 年）的平均收入水平。

2. 对 HANES 样本中 18－24 岁的男子，其身高与收缩血压之间的关系可概括如下：

   平均身高≈70 英寸，SD≈3 英寸

   平均血压≈124mm，SD≈14mm，r＝－0.2

   试估计 6 英尺高的男子的平均血压。

3. 某班学生期中考试平均分数为 60，SD 为 15，期末考试成绩亦如此。期中考试与期末考试成绩间的相关系数约为 0.50。估计期中考试取得如下分数的学生的期末考试成绩。

   (a)60　(b)75　(c)30

4. 对男孩子的跟踪研究得到如下结果：

   4 岁时平均身高≈41 英寸，SD≈1.5 英寸

   18 岁时平均身高≈70 英寸，SD≈2.5 英寸，r≈0.80

   （散点图见 155 页）试估计 4 岁时身高如下的男孩 18 岁时的平均身高

   (a)41 英寸　(b)44 英寸　(c)40 英寸

   如图 1 那样绘出你的回归估计。

5. 对 HANES 样本中 18－74 岁男子，有

   平均身高≈69 英寸，SD≈3 英寸

   平均体重≈171 磅，SD≈30 磅，r＝0.40

   试估计如下身高的男子的平均体重：

   (a)69 英寸　(b)66 英寸　(c)24 英寸　(d)0 英寸

   对(c)和(d)的答案进行评价。

6. HANES 样本（见习题 5）中 45－54 岁男子的平均身高为 69 英寸，与总平均身高相同。正确还是错误，并解释之：他们的平均体重应约为 171 磅。

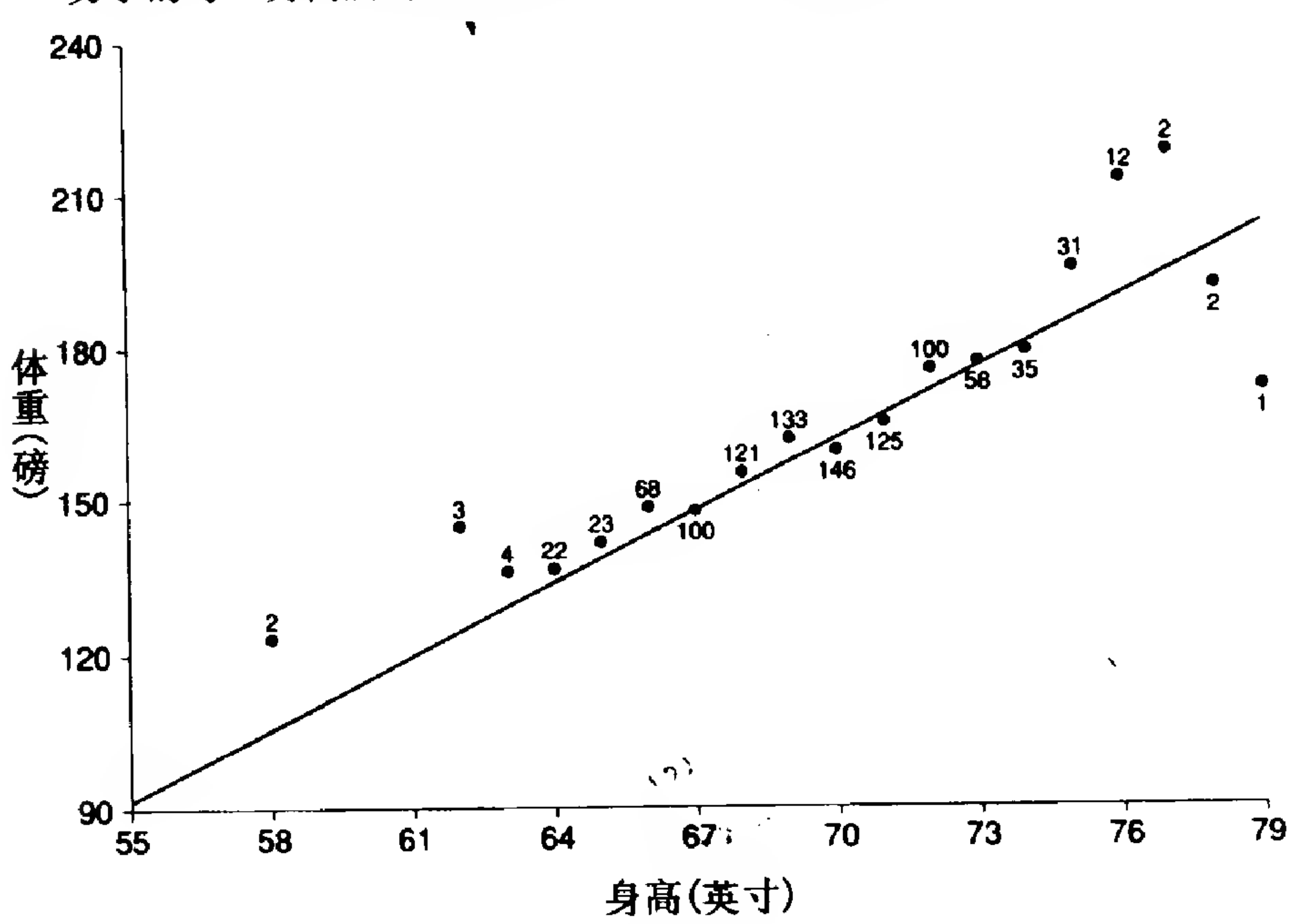
7. 假定 r＝－1，你能解释为什么 x 增加 1 个 SD 则相应的 y 减少 1 个 SD 吗？

这些习题的答案在第 688—689 页上。

2. 平均数图

图 4 是 HANES 样本中 18－24 岁男子身高和体重的平均数图[③]。例如，身高为 77 英寸（近似到英寸）的男子的平均体重为 218 磅，在图中用点（77 英寸，218 磅）表示。该组有 2 个男子，图中用点旁标的数字表示。

图 4 平均数点图。该图表示了 HANES 样本中 988 个 18－24 岁男子的每一身高所对应的平均体重。回归线将此图变得光滑

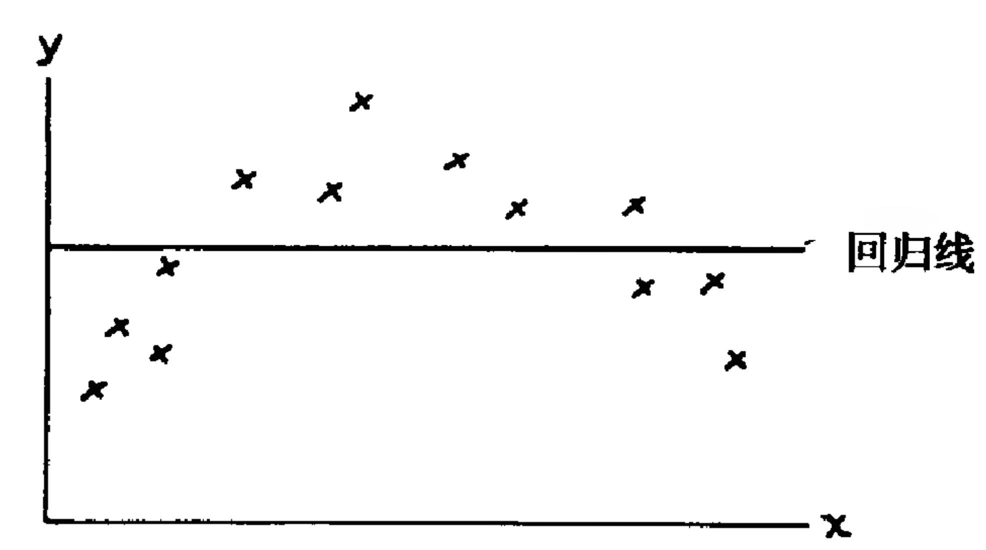


这种图把每一身高的男子的平均体重表示了出来。图的中部（多数人所处的位置）接近于一条直线，两端则起伏较大。例如，身高为 78 英寸的男子平均体重为 192 磅，这比身高为 77 英寸的男子的平均体重少 26 磅。这种情况下，身材高的男子反而比身材矮的男子体重轻。这是由于机会变异所致：这些样本的抽选是随机进行的，但碰巧抽中的男子中，77 英寸高的太重，78 英寸高的不够重。（当然，每个组都只有两人。）回归线则消除了这种机会变异。

回归线是平均数图的光滑形式。如果平均数图正好是一条直线，则回归线和平均数图必然合为同一条直线。

在有些情况，回归线消除得过多。如果两个变量之间存在非线性相关，如图 5，回归线将不注意这种情况，于是，最好使用平均数图。（对相关系数非线性也成问题——见第 9 章第 3 节。）

图 5 非线性相关。当变量间存在非线性相关时，不应该使用回归线。



习题 B

1. 下面有四个散点图，每个都有一条实线和一条虚线。对每一个图，说出哪一条是 SD 线，哪一条是 y 关于 x 的回归线。（见下页图）
2. 在一张纸上描出下图，将每一纵向条形的平均数点用一个叉标出（图中已标出了一个），然后绘出 y 关于 x 的回归线。（图中虚线为 SD 线）（见下页图）
3. 下面是一些虚拟的数据组。绘出每一组数据的散点图并画出平均数图。绘出 y 关于 x 的回归线。请不要进行任何计算，尽你所能进行推测。

(a)

x	y
1	3
2	5
3	7

(b)

x	y
1	0
1	6
2	5
3	6
3	8

(c)

x	y
0	0
0	2
1	2

(d)

x	y
0	0
1	1
2	4

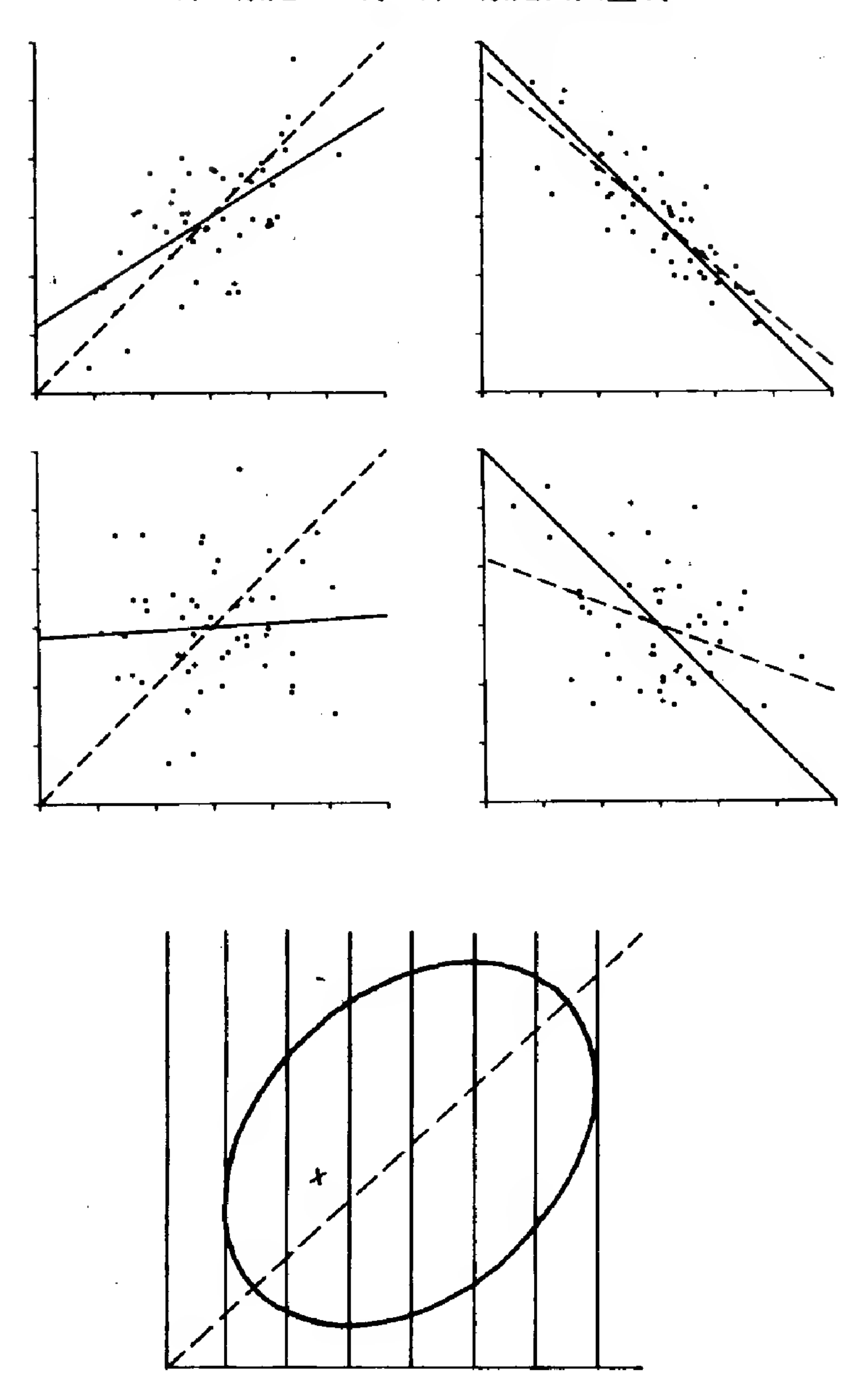
(e)

x	y
0	0
0	1
1	0
1	1

(f)

x	y
0	2
1	3
2	0
2	4
3	1
4	2

哪一条是SD线？哪一条是回归直线？



这些习题的答案在第 689 页上。

技术性注：一般情况下，按平均数图（每一平均数点按所代表的点的个数加权）拟合的回归线，与按原散点图拟合的回归线重合。（这在平均数图中 x

坐标不同的点互相分开的情况下是精确的；其它情况下，是一种较好的近似。）

3. 用于个体的回归方法

健康与营养检查调查中，18－24 岁男子身高与体重间的关系可概括如下：

平均身高≈70 英寸，SD≈3 英寸

平均体重≈162 磅，SD≈30 磅，r≈0.47

现在假设你必须猜测这些男子中某一个的体重而又不知道有关他的其他任何信息，最佳的猜测是总平均体重，162 磅。其次，告诉你该男子的身高：例如为 73 英寸。这位男子身材较高，体重可能比平均体重更重些。你对其体重的最佳猜测为该研究中所有 73 英寸高的男子的平均体重。这个新平均数可以用回归方法估计，为 176 磅。一般规则是：当你需从另一个变量预测一个变量时，采用新平均数。在许多情况，回归方法给出了估计新平均数的合理实用的方法。当然，如果变量之间存在着非线性关系，则回归方法不适用。

例 1. 某大学对完成第一学年学习的大学生的 SAT 数学分数（范围从 200 至 800）与第一年 GPA（范围从 0 至 4.0）之间的关系作了统计分析，结果为：

SAT 平均数＝550，SD＝80

第一年平均 GPA＝2.6，SD＝0.6，r＝0.4

散点图为橄榄球状。若一学生 SAT 分数为 650，预测其第一年的 GPA。

解　该生的 SAT 分数在平均数之上 100/80＝1.25SD。第一年 GPA 的回归估计为：在平均数之上 0.4×1.25＝0.5SD，即0.5×0.6＝0.3GPA 分。GPA 的预测值为：

2.6＋0.3＝2.9

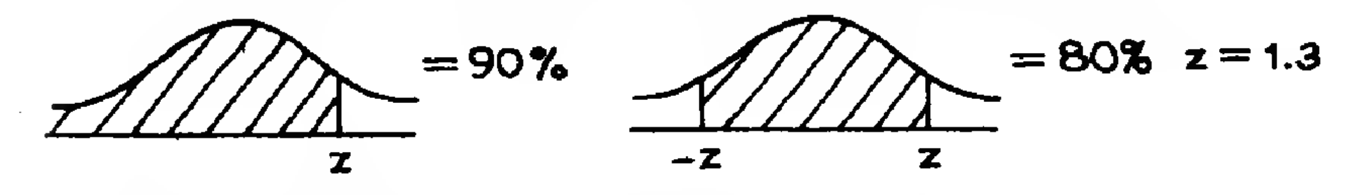
通常，调研人员通过研究求出回归估计，然后进行外推；他们将该估计用于新的对象。在许多情况，只要调查中的对象对即将被

作出推断的人具有代表性，这种方法都是有意义的。但是，你每次都必须考虑这个问题：回归方法中的数学原理并不保证你的结论的正确。在例 1 中，该大学只有关于它录取的学生的经验。把回归过程用于与之有很大差异的学生就可能会成问题。(最典型的是录取官员从录取的学生外推落选学生。)

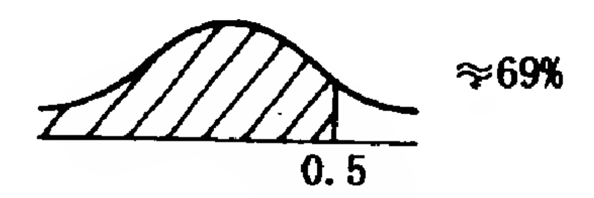
回到回归技术问题，回归技术也能用来预测百分位数排序。如果在测验中你的百分位数是 90%，说明你做得非常不错。班上只有 10%的学生分数比你高，其他 90%都比你低。25%的百分位数就不怎么样了：班上有 75%的学生分数比你高，只有其他 25%比你低[④]。

例 2. 例 1 中的大学发现 SAT 分数与第一年的 GPA 都服从正态分布。假定某个学生 SAT 的百分位数是 90%(在所有一年级学生中)，预测他第一年 GPA 的百分位数。

解　我们将采用回归方法，因而需要知道该学生的第一年 GPA 在平均数之上多少个 SD。他的百分位数隐含地给出了这一信息——因为 GPA 服从正态分布(见第 5 章第 5 节)：



这个学生的 SAT 在平均数之上 1.3SD。回归方法预测他的第一年 GPA 将在平均数之上 $0.4\times1.3\approx0.5$SD。最后，可以转换回到百分位数：



这就是答案。第一年 GPA 的百分位数排序预测为 69%。

在解这个问题的过程中，两个变量的平均数与 SD 值从来没有用到，所有要紧的是 r。基本地，这是因为整个问题是基于标准

单位进行的。百分位数排序给出的是标准单位，只不过是以规则的形式。

例 2 中的学生在 SAT 与第一年考试成绩这两个不同的方面与全班进行了比较。他的 SAT 非常不错，得分在第 90 个百分位数。但回归估计只把他第一年考试成绩放在第 69 百分位数上——仍高于平均数，但高得不太多。另一方面，对差的学生——如 SAT 的百分位数为 10% 的学生——回归方法的预测值将有所提高，在第一年的考试成绩中，他们将在第 31 百分位数上。这仍比平均数低，但接近了一些。

为了更仔细地着手于这事，取所有 SAT 成绩在第 90 百分位数的学生——成绩好的学生。他们中有的在第一年考试中名次将上升，有的将下降，但平均说来，这组人的名次将会下降。作为比较，再取所有 SAT 成绩在第 10 百分位数的学生——成绩差的学生。同样，他们中有的在第一年考试中名次将上升，有的将下降，但平均说来，这些人的名次将会上升。这就是回归方法给予我们的结果。

开始时许多人将预测第一学年的名次与 SAT 分数的名次相同。这不是一种好策略。为了明了这一点，设想你需预测一学生在数学班的名次。在缺乏其他信息的情况下，最稳当的猜测是将她放在中位数。但是，如果你知道该生的物理成绩很好的话，你会猜她的数学名次在中位数之上，毕竟在物理和数学之间存在强相关。另一方面，如果所有你知道的是她在陶瓷班的名次，则对猜测她的数学名次毫无帮助：回到中位数。陶瓷和数学间没有什么相关。

现在，回到从 SAT 名次猜测第一学年名次的问题上。如果这两种分数完全相关，则预测第一学年名次与 SAT 名次相同是讲得通的。在另一种极端，如果两者间是零相关，则 SAT 名次对预测第一学年名次毫无帮助。事实上，这种相关处于两个极端之间某个地方，因此，你只得预测第一学年考试分数的名次在 SAT 名次与中位数之间的某个地方。

习题 C

1. 某班期中考试平均分数为 60,SD 为 15,期末考试分数也如此,期中考试分数与期末考试分数间的相关系数约为 0.50。预测期中考试分数如下的学生的期末考试分数:

   (a)未知　(b)60　(c)75　(d)30

   并将你的结果与 P183 的习题 3 比较。

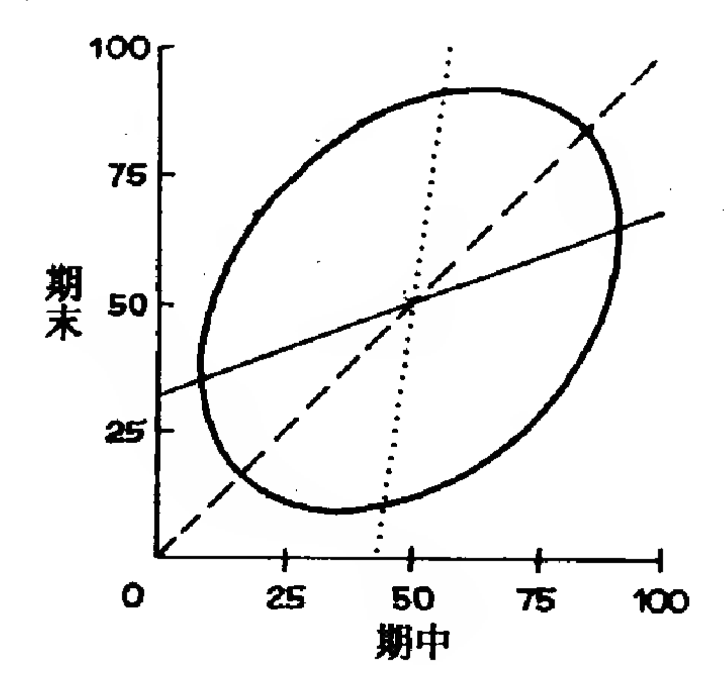
2. 某大学第一学年学生 SAT 分数与第一学年 GPA 间的相关系数为 0.60,预测 SAT 百分位数如下的学生第一学年 GPA 的百分位数:

   (a)未知　(b)50%　(c)90%　(d)30%

   并将(c)的答案与例 2 进行比较。

3. HANES 样本中 18-74 岁男子身高与体重之间的相关系数为 0.4;身高的 SD 约为 3 英寸。65-74 岁男子平均比 25-34 岁男子的身高矮约 2 英寸。正确还是错误并解释之:因为 2 英寸为身高的 0.67 个 SD,则 65-74 岁男子比 25-34 岁男子的平均轻约 04×0.67≈0.27 个体重 SD。

4. 下面的散点图给出某课程的期中和期末考试分数。有三条直线穿过该图,在两次考试中的百分位数都相同的学生被画在其中之某一条直线上,问是哪一条,为什么?



这些习题的答案在第 689—690 页上。

技术性注:例 2 讨论的方法是中位数排序法。假设正态性且 r=0.4,SAT 百分位数为 90%(相对于同班同学)的学生中,约一半的第一学年 GPA 的百分位数在 69%之上,另一半将在其之下。估计平均排序的过程要更复杂一

些。

4. 回归谬论

某学前班企图增加儿童的智商。儿童进入学前班时接受一次测试(学前测试),结业时再接受一次测试(学后测试)。两种情况下的分数平均接近于100,SD约为15。由此看来学前班似乎没什么效果。但进一步分析数据,可得出一些令人吃惊的结果。在学前测试中低于平均数的儿童在学后测试中平均提高了约5个智商点。与此相反,在学前测试中高于平均数的儿童,在学后测试中平均降低了约5个智商点。这说明了什么呢?学前班的作用就是将智力平均化吗?也许较聪明的儿童与较笨的儿童在一起玩时,两者之间的差异趋于消失;这种说法是对还是错?

上述推测也许很有趣,但实际上好事坏事都未发生。给出一个理由,不可能期望儿童在两次测试中的分数完全一致。两次测试的分数将会有差异。没有人会把这些差异看得很重,或者需要任何解释。但这些差异使测试结果的散点图围绕SD线散布并形成熟悉的橄榄球状云图。围绕直线的散布使下面的部分上升,上面的部分下降。没有其他原因。

> 事实上,在所有考试—再考试的情形中,在第一次考试中最低的那部分,在第二次考试中将平均有所提高——而最高的部分将平均有所降低。这就是回归效应。

回归谬论认为回归效应是由于某些重要因素引起,而不仅是围绕直线的散布所致。

我们来看看为什么只要有围绕SD线的散布就会产生回归效应的原因。Galton在他关于家庭成员相似性的研究中首先注意到这种效应,因此,我们也以这项研究来进行讨论。但是这种推理是普遍性的。图6给出第八章中讨论过的1 078对父子身高的散点图,概括统计量为[5]:

父亲平均身高≈68 英寸,SD≈2.7 英寸

儿子平均身高≈69 英寸,SD≈2.7 英寸,r≈0.5

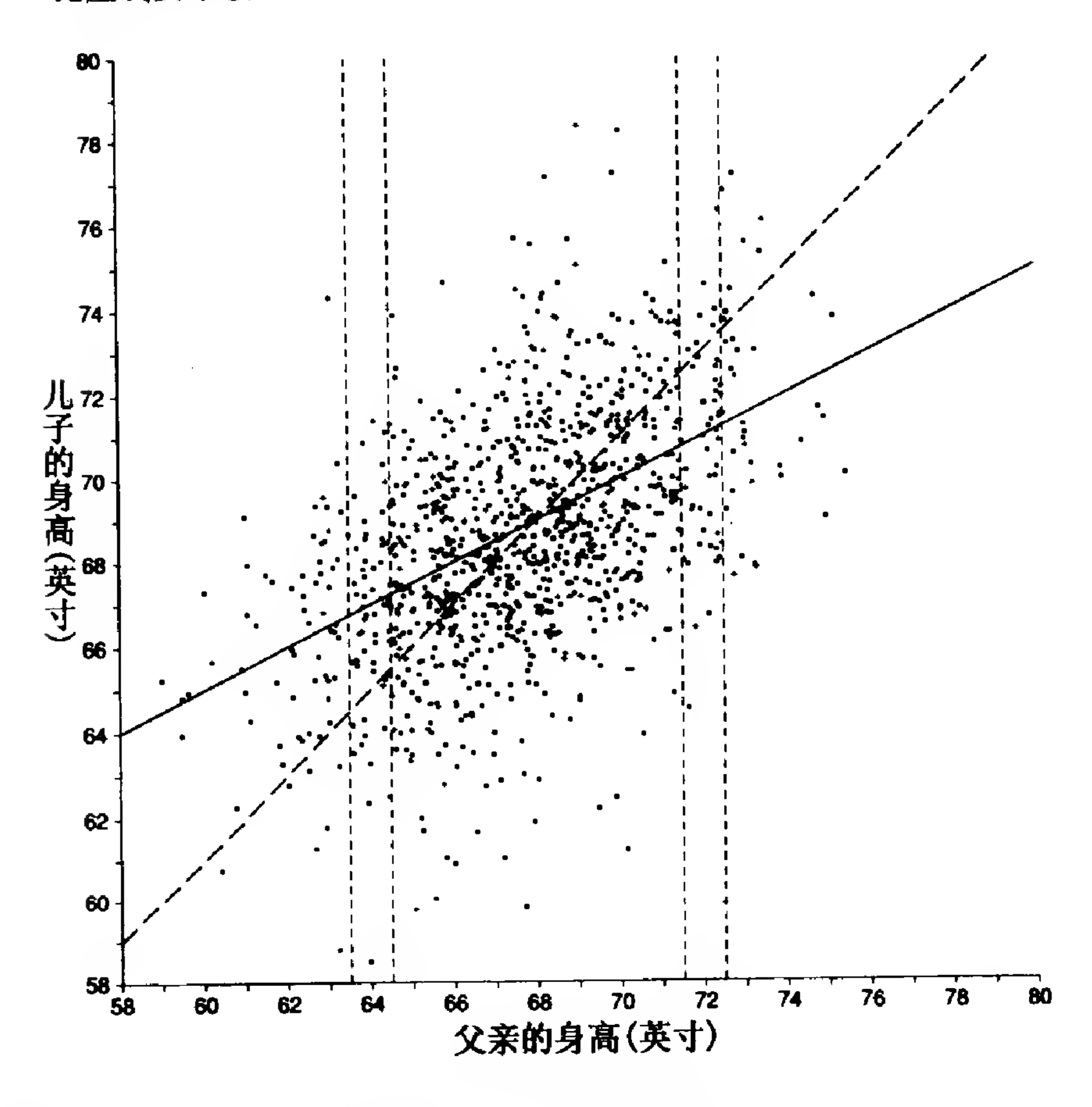
儿子平均比父亲高 1 英寸。由此,很自然地会猜测 72 英寸高的父亲会有 73 英寸高的儿子;类似地,64 英寸高的父亲会有 65 英寸高的儿子,如此等等。这样的父亲和儿子在图 6 中沿虚线表示。当然,恰好在这条直线上的家庭并不太多。事实上大量的家庭散布在直线两旁。有的儿子比父亲高,有的比父亲矮。

例如,以身高为 72 英寸(近似到英寸)的父亲为例,相应的家庭标绘在图 6 中 72 英寸上方的纵条,儿子的身高有相当的范围。有的点在虚线上方:这些家庭中儿子身高高于 73 英寸,但大多数点在虚线下方:这些家庭中儿子身高比 73 英寸矮。总的来说,72 英寸高的父亲其儿子的平均身高为 71 英寸。身材高的父亲(第一次考试中的高分),其儿子的平均身高将下降(第二次考试中分数下降)。

现在来看看 64 英寸上方的纵条,它代表所有父亲身高为 64 英寸(近似到英寸)的家庭。那里虚线的高度为 65 英寸,代表比他那 64 英寸高的父亲高 1 英寸的儿子。有的点落在虚线之下,但大多数点在虚线之上,64 英寸高的父亲其儿子的平均身高为 67 英寸。对于身材矮的父亲(第一次考试中的低分),平均而言其儿子高一些(第二次考试中分数上升)。高贵的 Galton 把这称作"回归到平常"。

图 6 中的虚线穿过对应于父亲平均身高 68 英寸和儿子平均身高 69 英寸的点。沿着虚线,父亲身高每增加 1 个 SD,相应地儿子身高也增加 1 个 SD。这两个事实使它就是 SD 线。点云图关于 SD 线对称,但 72 英寸上的纵条关于 SD 线不对称。该纵条只包含 x 坐标不同寻常地大的点,纵条中的大多数点落在 SD 线之下。与此相反,64 英寸上的纵条只包含 x 坐标不同寻常地小的点,该纵条中大多数点落在 SD 线之上。这种内在的不平衡在橄榄球状点云图中总是存在的。回归效应的这种图形解释似乎不太浪漫,但

图6 回归效应。散点图表示1 078对父亲和儿子的身高。儿子平均比父亲高1英寸。儿子比父亲高1英寸的家庭画在虚线上。72英寸上方的条形图中的点对应于所有父亲身高为72英寸(近似到英寸)的家庭,这些点的大多数在虚线之下,其儿子的平均身高只有71英寸。64英寸上方的条形图中的点对应于所有父亲身高为64英寸(近似到英寸)的家庭,这些点的大多数在虚线之上,儿子的平均身高为67英寸。实回归线穿过所有纵向条形图的中心,比虚线要平缓一些。



话说回来,统计学并不是富于浪漫色彩的学科。

图6同时也给出了儿子身高关于父亲身高的回归线。这是一条比虚的SD线上升较缓的实线,它穿过每一圆点上x纵条的中

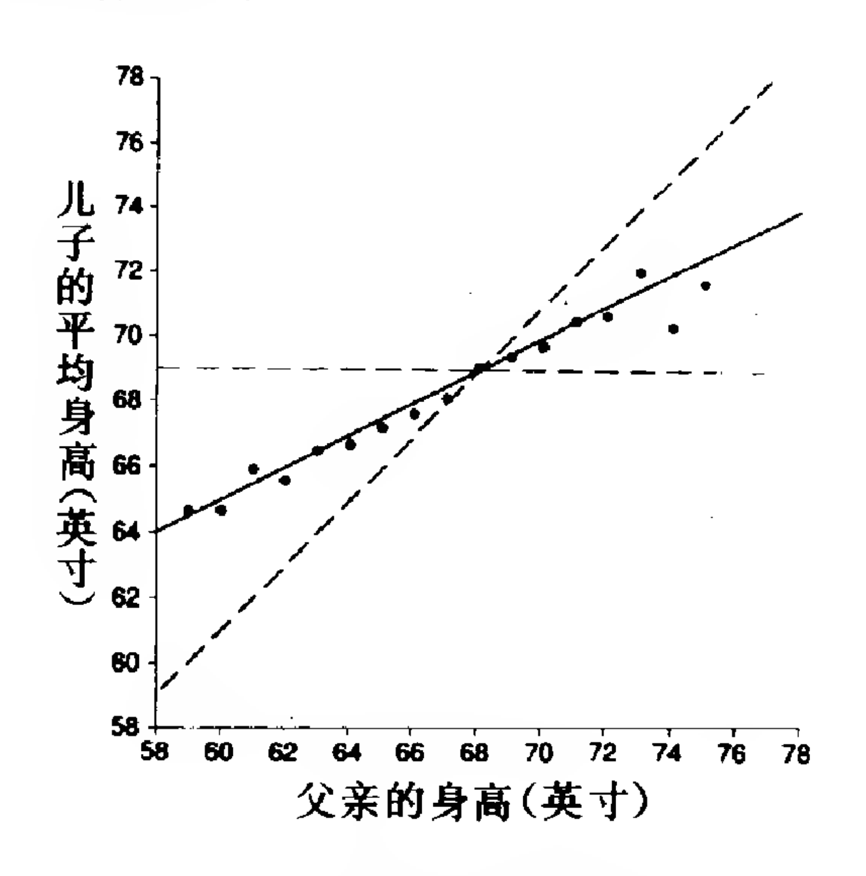
心——纵条中点的平均 y 值。例如，取身高 72 英寸的父亲，他们比平均数高 4 英寸；4 英寸/2.7 英寸≈1.5SD。回归线指出他们的儿子将高于平均数大约

$$r \times 1.5SD = 0.75SD \approx 2\text{ 英寸}$$

儿子的总平均身高是 69 英寸，因此他们的儿子平均身高的回归估计为 71 英寸——正对上。

图 7 不用点云图，采用最明显的方式显示回归效应。虚的 SD 线以 45 度角上升。圆点表示了对应于每一个父亲身高值的儿子身高的平均数。这些圆点是图 6 中纵条的中心，圆点的上升比 SD 线平缓。这就是回归效应。总体上，圆点处于 SD 线与穿过

图 7 回归效应。SD 线是虚的，回归线是实的。圆点表示了对应于每一父亲身高值的儿子平均身高，它们的上升不如 SD 线那样陡。这就是回归效应。回归线与圆点的确拟合得很好。



平均数点的水平线之间的当中。这是因为相关系数是 0.5，因此父亲身高每增加 1 个 SD，儿子身高的增加是 0.5SD，而不是 1 个 SD。实的回归线以 0.5 比 1 的比例上升，与平均数图拟合得确实非常好。

第一眼就看到图 6 中，散点图是非常混乱的。Galton 天才地从混乱中发现了一条直线。Galton 时代以来，许多其他研究者都发现在他们的散点图中的平均数点也是非常接近于直线的。这就是为什么回归线非常有用的原因。



现在，来看看现象的背后：有些场合回归效应稍许容易被理解，例如上文提到的重复智商测试。基本事实是两次测试的分数倾向于有较大的差异，这可以解释为机会变异。每个人在第一次测试中可能运气较好，也可能较差。但如果某人第一次测试的分数很高，说明他在这次测试中运气很好，这就意味着他在第二次测试中的分数很可能会低一些。（你不会说："他的分数很高，他那天的运气真糟。"）另一方面，如果在第一次测试中分数很低，说明此人在那次不走运到某种程度，下次会好些。

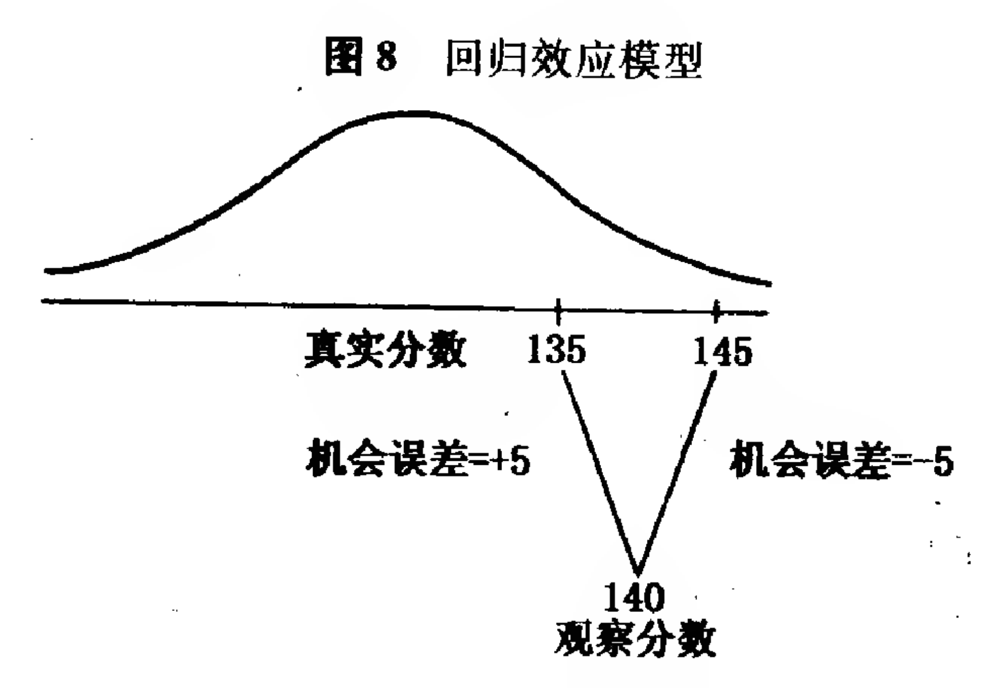
下面是考试—重考试问题的原始模型，由此可进一步解释回归效应。基本方程式为：

考试的观察分数＝真实分数＋机会误差

假定该总体的真实分数服从平均数为 100、SD 为 15 的正态分布。

再假定机会误差为正值和负值的可能性一样，大小约为 5 分。一个真实分数为 135 的人，在测验中可能得分 130 或 140。一个真实分数为 145 的人，在测验中可能得分 140 或 150。（当然，机会误差也可能是 ± 4，或 ± 6，等等；任何对称值可以用类似的方法处理。）

图 8　回归效应模型



对于第一次考试分数为 140 的人来说，此观察分可能有两个解释：

- 真实分数低于 140，连着一个正的机会误差；
- 真实分数高于 140，连着一个负的机会误差。

第一种解释更可能些：例如，如图 8 所示，真实分数为 135 的人比 145 的人要多些。

模型说明了回归效应的原因：如果某人第一次考试的分数高于平均数，他的真实分数可能比观察值略低。如果此人再次参加考试，我们预测第二次考试分数将比第一次考试分数低一些。另一方面，如果某人第一次考试的分数低于平均数。我们估计他的真实分数比此观察分略高一些，从而对第二次考试的分数的预测将略高于第一次考试分数。

习题 D

1. 作为训练的一部分，空军飞行员将与教官一起进行两次着陆实习，并对其

表现作出评定。在每次着陆后教官都与飞行员讨论评定成绩。统计分析表明，第一次着陆表现较差的第二次表现就较好。反过来，第一次着陆好的第二次就较差。由此得出结论：批评使飞行员进步，表扬使他们退步。因而，教官受命批评所有的着陆，不管是好还是坏。问事实是否支持这种做法？回答是或否。并简短地解释。

2. 某教师把期中和期末考试成绩进行了标准化，使全班两次考试的平均分数都是 50，SD 都为 10。两次考试成绩的相关系数总是约为 0.50。有一次，她把所有期中考试分数在 30 以下的学生列出并给予特别指导，结果他们的期末考试成绩都在 50 以上。这能用回归效应进行解释吗？回答是或否，并简短地解释。

3. 在美国教育分布中处于第 90 百分位数的人中，其收入分布在 90% 分数位以上的人所占的百分比为：

   大约 50%　50%以上　50%以下　不能确定

   选择一项并简短地解释。

4. 在图 6 和图 7 的数据中，61 英寸高的父亲的儿子平均比 62 英寸高的父亲的儿子高，还是矮？如何解释？

5. 某学生写道："回归效应就是事情趋于平均。"在图 6 中，父亲的总平均身高为 68 英寸，儿子的为 69 英寸。取父亲身高为 68 英寸（近似到英寸）的家庭——即 68 英寸上方的纵条，用下列数字填空：

   1/4　1/2　1　2　4　68　69　70　71

   (a) 在此纵条中，父亲身高约为____英寸。

   (b) 在此纵条中，儿子身高为____英寸。±____英寸。

   (c) 简短地评价引语。什么"事情"趋于平均，什么"事情"不是这样？

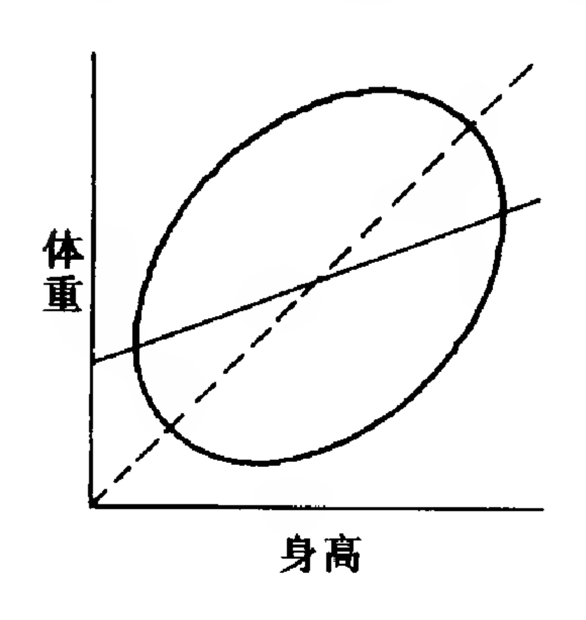
这些习题的答案在第 690—692 页上。

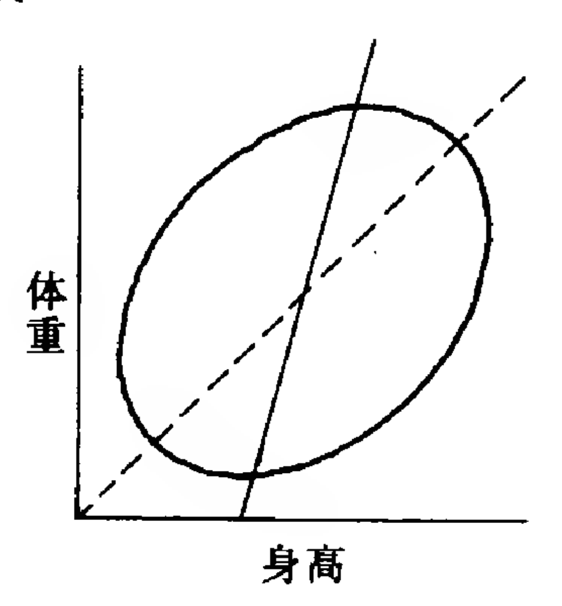
5. 有两条回归线

事实上，穿过散点图可画出两条回归线。例如，下面的图 9 画出了身高与体重的散点图。左图表示的是体重关于身高的回归线，它穿过所有纵条的中心，估计了对应于每一身高的平均体重。右图表示的是身高关于体重的回归线，它穿过所有横条的中心，估计了对应于每一体重的平均身高。在两图中，回归线用实线，SD 线用

虚线表示。为多数目的，似乎体重对身高的回归更自然，但是，另一条也可能很方便。

图9　身高与体重。左图表示体重对于身高的回归线；右图表示身高对于体重的回归线。虚线是SD线。





例3. 对智商分进行适当调整，使男子和女子的平均智商都约为100，SD约为15。丈夫和妻子智商之间的相关系数约为0.50。一项大型家庭研究发现，智商为140的男子，其妻子的平均智商为120。考虑此研究中智商为120的所有妻子，他们的丈夫的平均智商比120高吗？回答是或否，并简短地说明。

解：否。她们丈夫智商的平均数为约110。见图10，丈夫智商为140的家庭表示在相应的纵条中，此纵条点的y的平均数为120。妻子智商为120的家庭表示在横条中，这是一个完全不同的家庭族。横条中点的平均x值约为110。记住，有两条回归线：一条是由丈夫的智商预测妻子的智商；另一条是由妻子的智商预测丈夫的智商。

习题E

1. 对HANES样本中18—24岁男子来说，身高为73英寸者的平均体重约为176磅。正确还是错误并解释之：体重为176磅者的平均身高必约为73英寸。

图 10　两条回归线

2. 在 Pearson 的研究中，身高为 72 英寸父亲的儿子的平均身高为 71 英寸。正确还是错误：如果儿子身高为 71 英寸，他们的父亲的身高平均将约为 72 英寸。请简短地解释。
3. 在例 2（第 188 页）中，回归方法预测 SAT 百分位数在 90%的学生在第一学年中 GPA 的百分位数仅为 69%。正确还是错误并解释之：第一学年 GPA 的百分位数为 69%的学生的 SAT 百分位数为 90%。

这些习题的答案在第 692 页上。

6. 复习题

复习题可能包含前面各章的内容。

1. Pearson 和 Lee 在对约 1 000 个家庭所进行的研究中获得了如下结果：

   丈夫平均身高≈68 英寸，SD≈2.7 英寸

   妻子平均身高≈63 英寸，SD≈2.5 英寸，r≈0.25

   当丈夫身高如下时预测妻子的身高：

   (a)未知　　(b)68 英寸　　(c)72 英寸　　(d)64 英寸

2. 在对智商稳定性所进行的研究中，一大群人在 18 岁时被检测一次，35 岁时又检测一次，获得如下结果：

   18 岁：平均分≈100，SD≈15

   35 岁：平均分≈100，SD≈15，r≈0.80

(a)估计 18 岁时智商分为 115 的所有人在 35 岁时的平均智商分。

(b)某人 18 岁时智商分为 115，预测其 35 岁时的智商分。

3. 在一项研究中，居住于某城镇的丈夫和妻子的受教育水平间的相关系数为约 0.50，他们的平均受正规学校教育年数都为 12 年，SD 为 3 年[⑥]。

(a)某妇女的丈夫受正规学校教育年数为 18 年，预测她的受教育水平。

(b)某男子的妻子受正规学校教育年数为 15 年，预测他的受教育水平。

(c)很明显，受教育程度高的男子娶受教育程度比他们低的女子为妻。但妇女嫁给所受的教育程度比她们还低的男子，这是否可能？

4. 某医生习惯测量两次血压。他注意到第一次测量时血压不寻常地高的病人第二次测量时有所降低。他的结论是第二次测量时病人比较放松。一同事不同意，指出第一次测量时血压不同寻常地低的病人第二次测量时有所上升，暗示了病人更紧张了些。哪个医生的意见正确？或也许两个医生的结论都不对？简短地解释之。

5. 在前一题已涉及的关于血压问题的一项大型研究中，发现第一次测量时平均读数为 130mm，第二次测量时平均读数为 120mm。这支持了哪位医生的论点？或这是回归效应吗？请说明。

6. 在一个大统计班中，每学期期中考试和期末考试分数间的相关系数接近 0.50。散点图为橄榄球状。预测期中考试分数的百分位数如下的学生在期末考试中的百分位数。

(a)未知　(b)50%　(c)5%　(d)80%

7. 正确还是错误：在美国，身高在身高分布中的百分位数为 40% 的人，其体重在体重分布中的百分位数也为 40%。并简短地说

明。

8. 某研究人员测量了一大群运动员的各种特性，发现运动员体重与他能举起的重量之间的相关系数为 0.60。正确还是错误并解释之：

(a)平均来说运动员能举起其体重的 60%的重量。

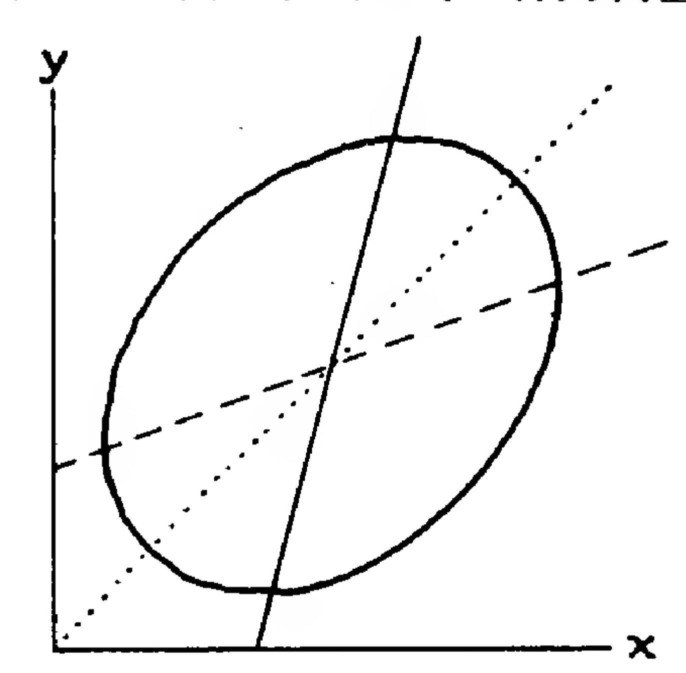
(b)如果运动员的重量增加 10 磅，他可指望多举 6 磅的重量。

(c)运动员的体重越重，他平均能举起的重量越多。

(d)运动员能举起的重量越多，他的平均体重越重。

(e)运动员举重能力的 60%可归因于其体重。

9. 穿过下面的散点图，画有三条直线，一条是 SD 线，一条是 y 对 x 的回归线，一条是 x 对 y 的回归线。各为何直线，并说明理由。



7. 小结

1. 与 x 每增加 1 个 SD 相应，y 平均只增加 r 个 SD。画出这些回归估计可得 y 对 x 的回归线。

2. 平均数点图常接近于一条直线，只是可能略有波动。回归线消除了这些波动。如果平均数点图是一条直线，则它与回归线相重合。如果平均数点图呈强非线性形式，此时回归不适用。

3. 回归直线能用于预测个体值。但如果需外推的数据较远，或

者是属于不同类的问题时，需特别当心。

4. 在典型的考试一重考试情形中，被测对象在两次考试中得到不同的分数。对在第一次考试中分数最低的群体来说，在第二次考试中的分数有的升高，有的降低；但平均起来，分数最低的群体的分数将有所上升。对第一次考试中分数最高的群体来说，在第二次考试中的分数有的将上升，有的将下降；但平均起来，分数最高的群体的分数将下降。这就是回归效应。只要散点图围绕 SD 线的分布呈橄榄球云图，回归效应就会出现。

5. 回归谬论认为，回归效应归因于围绕 SD 线的散布以外的其他原因。

6. 在同一个散点图中可以画出两条回归线：一条由 x 预测 y；另一条由 y 预测 x。

11 回归的均方根误差

这些是正态相关的数学上的形式推论。许多生物统计学资料肯定地显示与这种设想所预期的特性大体一致:尽管我不知道这个问题受到过任何充分严格的调查。近似一致可能完全足以证实相关用作描述总体的一个量是合理的;它在这方面的功效是无可置疑的,而且这不是不可能的,即在某些场合,它连同均值和方差对变量的联立变差提出了一个完整的描述。

——R. A. Fisher 爵士(英格兰,1890—1962)①

1. 引言

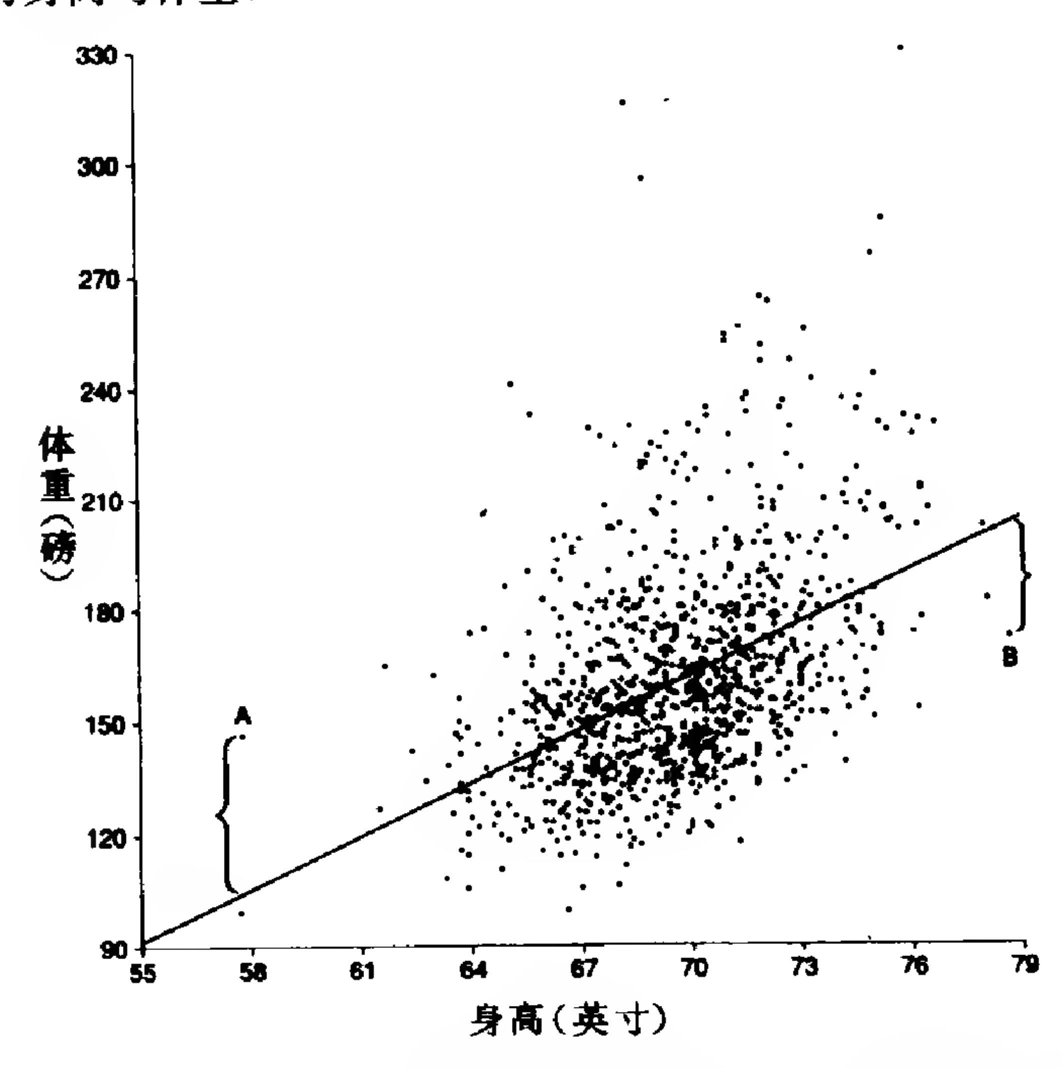
回归方法可用来从 x 值预测 y 值。然而,实际值与预测值常常是不一样的。它们之间的差异有多大呢? 本节的目的是用均方根误差度量这种差异的总体大小。以 HANES 样本中 988 个 18—24 岁男子的身高与体重为例(第 10 章第 1 节)。概括统计量为:

平均身高≈70 英寸,SD≈3 英寸

平均体重≈162 磅,SD≈30 磅,r≈0.47

简短地回顾一下,已知某人的身高,则他的体重可用具有同样身高的所有男子的平均体重来进行预测。这个平均体重可用回归方法估计;这种估计值都将落在回归直线上(见图 1)。图中的 A 近 58 英寸高,对应于此身高的平均体重的回归估计为 106 磅(见第

图 1 残差。残差是点在回归直线之上(+)或之下(-)的距离(距离是按垂直方向测量的。)。每一距离都对应于回归直线的一个预测误差。此散点图代表 HANES 样本中 988 个年龄在 18-24 岁男子的身高与体重。



10 章第 1 节)。然而,A 的实际体重为 146 磅。预测值与实际值不等,误差或残差为 40 磅:

残差=实际重量-预测重量

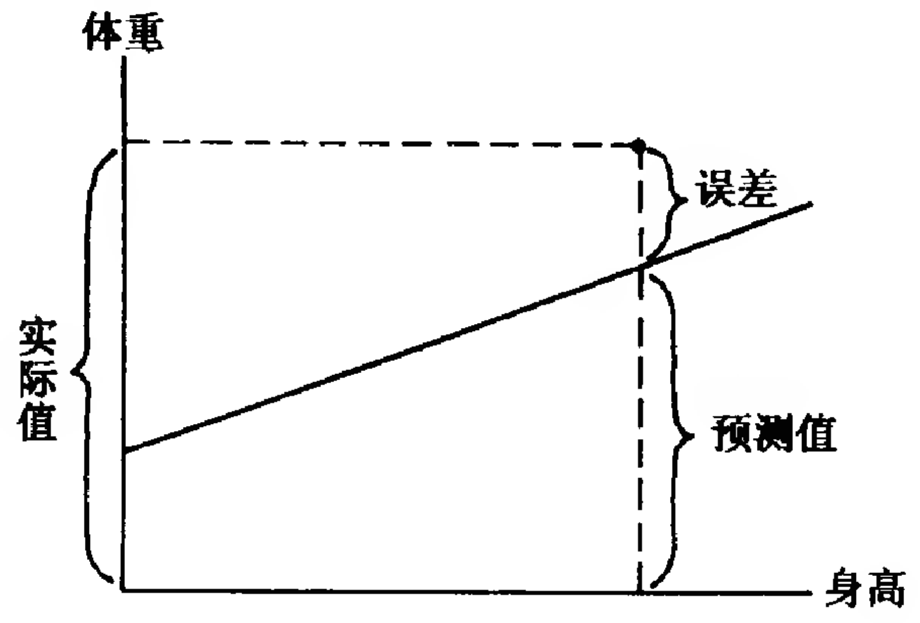
　　=146 磅-106 磅=40 磅

图中,该预测误差是 A 在回归直线上方的纵向距离,用大括弧标记。

图中的B身高约79英寸，体重172磅。用回归线预测其体重为204磅。因此，预测误差为172磅－204磅＝－32磅。在图中，此残差为B在回归直线之下的纵向距离，仍用大括弧标记。

某一点在回归直线之上（＋）或之下（－）的距离为 　　残差＝实际值－预测值 此残差是用回归方法进行预测的预测误差的图形表示。

图2　预测误差等于点与直线的纵向距离



散点图中每一个点都有一个残差，代表回归方法带来的误差。图2说明了误差与点距直线的距离之间的关系。这些误差的总体大小可用它们的均方根进行度量（见73页），结果称为回归直线的均方根误差。用散点图来表示，回归直线的均方根误差能告诉你一个代表点距回归线有多远。图3对此进行了说明。

散点图中的点按与均方根误差大小相似的残差偏离回归直线（上或下）。均方根误差与回归直线间的关系相当于标准差与平均数间的关系。

散点图中约有68%的点将在与回归直线相距一个均方根误差的范围中；约有95%的点将在与回归线相距两个均方根误差的范围中。这种约略估计对多数数据来说是正确的，但不是对所有数据。在图4中得以说明。

图 3 均方根误差。它说出了代表点在回归线之上或之下的距离。

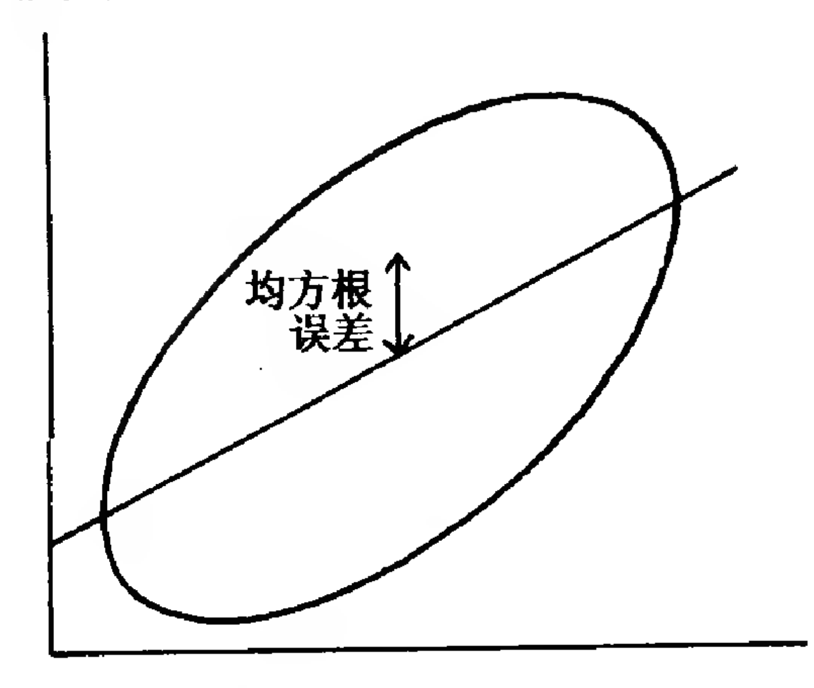
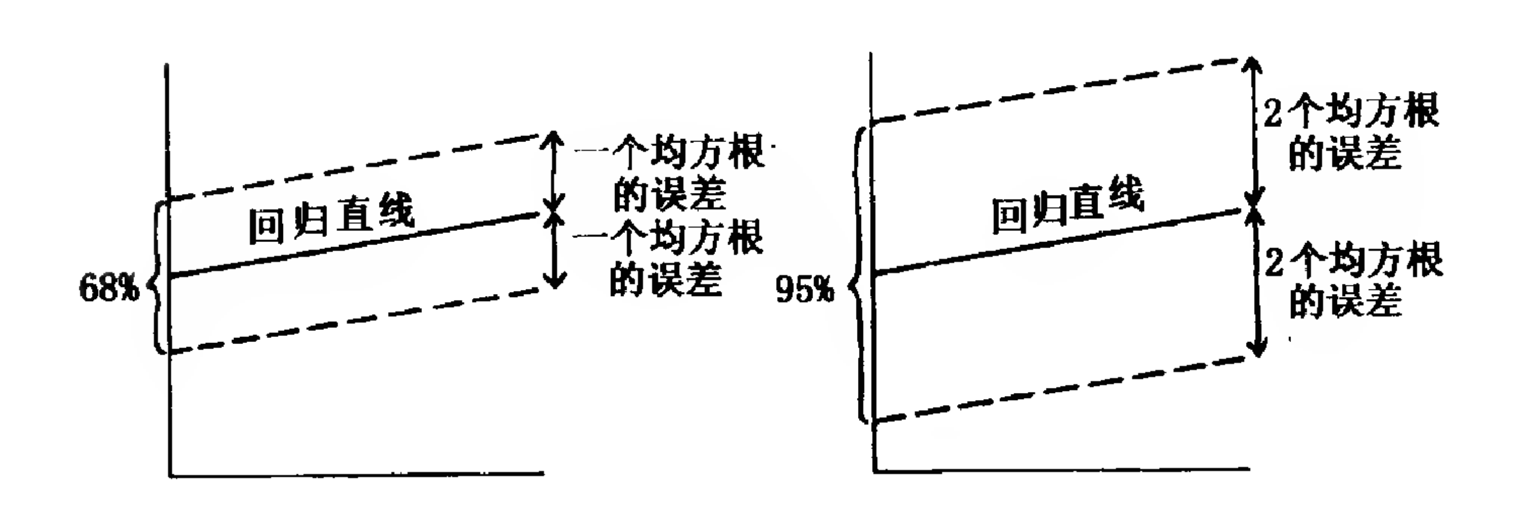


图 4 均方根误差。散点图中约 68%的点落在边界与回归线平行的条形图中，边界与回归线的距离为一个均方根误差（上或下）。约 95%的点落在边界与回归线平行的更宽的条形图中，边界与回归线的距离为二个均方根误差。

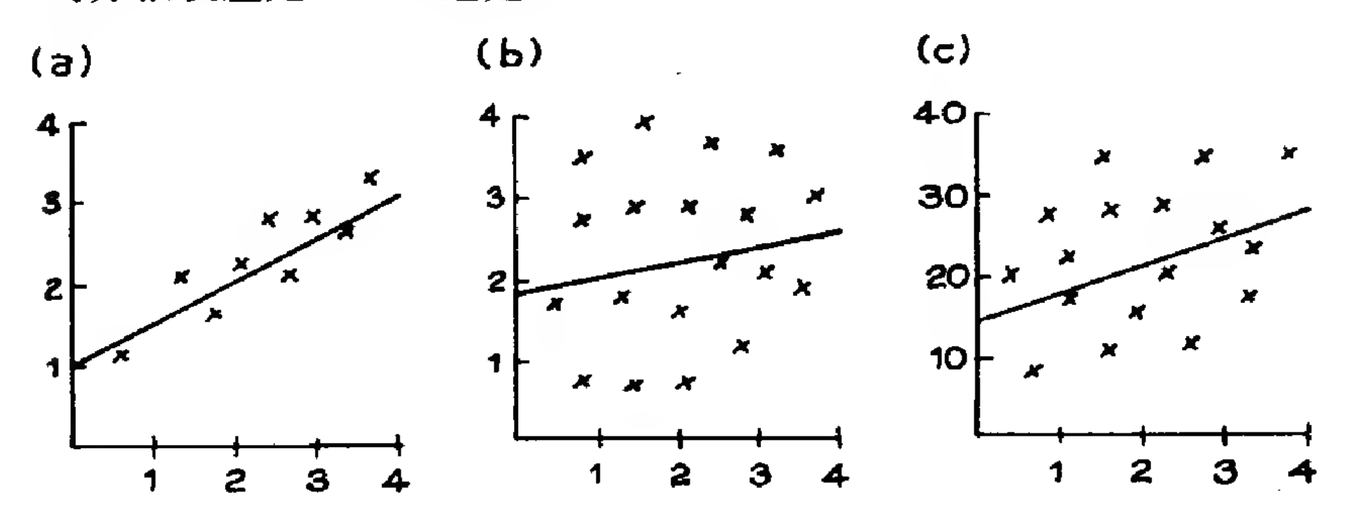


在图 1 中，均方根误差为 26 磅。对此的图形解释是：代表点在回归直线上或下的距离约为 26 磅。由于这条直线是从身高预测体重，我们得出的结论是：这项研究中代表性男子的实际体重与预测体重间的差异约为 26 磅左右。68%与 95%规律如何呢？通过计算机发现，988 个男子中 937 人在两个均方根误差（52 磅）的范围中，亦即 95%。这里，约略估计相当不错。另一方面，988 个男子中 752 人的预测值正确到一个均方根误差（26 磅）的范围内——76%。

68%是一个粗糙的近似值。

习题 A

1. 下面有三个散点图。用目测方法画出的回归直线穿过散点图。猜测各图中均方根误差是 0.2、1 还是 5。



2. 预测收入的回归直线具有均方根误差为 2 000 美元。用该线预测某人的收入为 20 000 美元。这个预测在多大范围内正确:几百美元、几千美元、一万或二万美元?
3. 一招生人员试图在两种预测第一学年分数的方法中选择一种。一种方法的均方根误差为 12,另一方法的均方根误差为 7,其他条件相同。他应该选哪一种方法,为什么?
4. 预测考试分数的回归直线具有均方根误差为 8 分。
   (a)此时 68%的预测值将在____分之内是正确的。
   (b)此时 95%的预测值将在____分之内是正确的。

这些习题答案在第 693 页上。

2. 均方根误差的计算

从数据直接计算均方根误差需用较长的时间。这里有一种简捷计算法:

> y 关于 x 的回归直线的均方根误差可用下式计算:
>
> $\sqrt{1-r^2}\times$y 的 SD

式中取哪一个变量的SD？应该采用待预测变量的SD。如果由身高预测体重，应取体重的SD，均方根误差将以磅为单位，不是英寸。如果由受教育水平预测收入，应取收入的SD，均方根误差将以美元为单位，而不是年。

均方根误差的单位与待预测变量的单位相同。

在第一节的身高体重的例子中，用回归从身高预测体重的均方根误差为：

$$\sqrt{1-r^2}\times \text{体重的 SD}=\sqrt{1-(0.47)^2}\times 30\ \text{磅}\approx 26\ \text{磅}$$

这个均方根误差并不比体重的SD小多少，这是因为体重与身高的相关关系不够密切：$r\approx 0.47$。知道某人的身高对预测其体重没有太大的帮助。

更一般地，上述公式把两件事联系起来：回归直线的均方根误差和y的SD。这些统计量描述了散点图的不同的方面。首先来看y的SD，图5中的左图显示了一条穿过y的平均数的水平线。y的SD度量的是一个代表点在该水平线之上或之下的距离。也就是说，如果你不考虑x，用y的总平均数来预测y，则SD度量了误差的大致大小。作为对比，均方根误差度量了代表点在回归直线之上或之下的距离（右图）。均方根误差要小一些，是因为回归直线（它可以按变化趋势向上或向下倾斜）更接近于散点。上述公式给出了均方根误差比SD小的精确因子：$\sqrt{1-r^2}$。

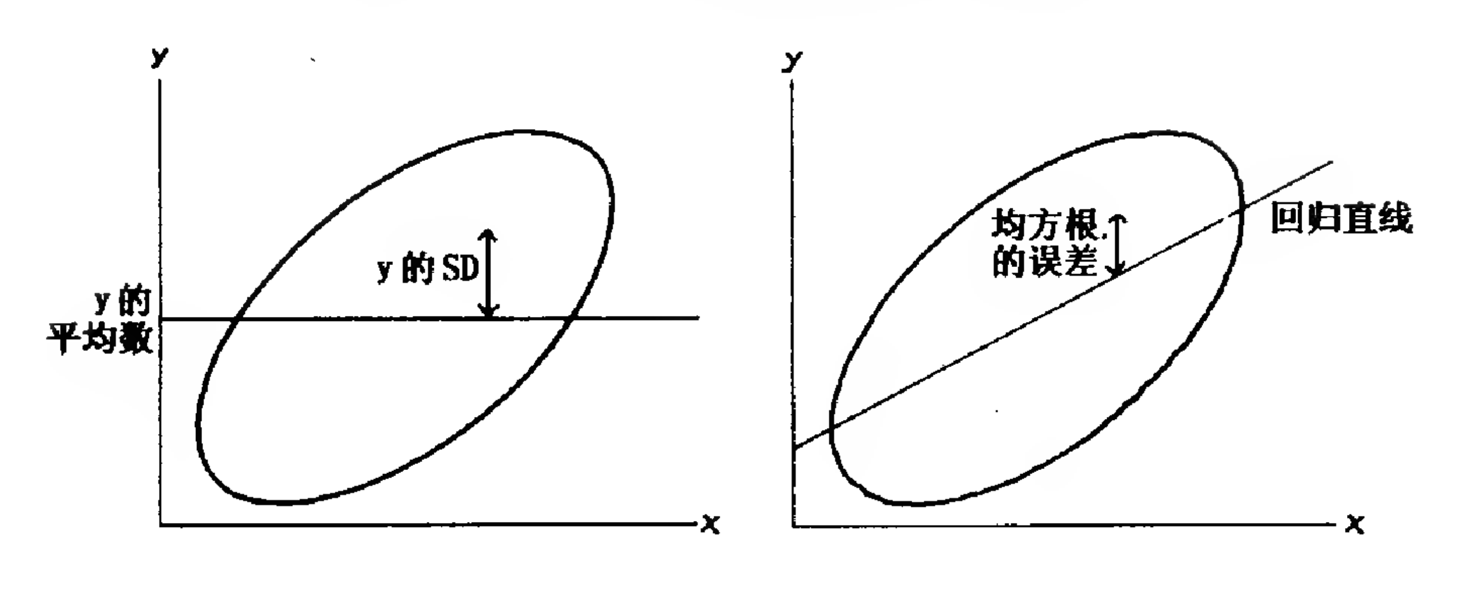
均方根误差的计算公式不用代数很难予以证明，但有三种特殊情况很容易理解。第一种情况，假定$r=1$，所有点都在一条直线上（直线斜向上），回归线穿过散点图中的所有点，所有的残差都为0。因此，均方根误差也是0。这与公式相符：因子等于

$$\sqrt{1-r^2}=\sqrt{1-1^2}=\sqrt{1-1}=0$$

$r=-1$的情况是一样的，不同的只是直线是斜向下的。均方根误

差仍将是 0,与公式相符:

图 5　y 的 SD 与回归直线的均方根误差



$$\sqrt{1-r^2}=\sqrt{1-(-1)^2}=\sqrt{1-1}=0$$

第三种情况是 r=0,此时变量间没有线性关系。因此回归直线对预测 y 值毫无用处,均方根误差将等于 SD。公式给出的因子是:

$$\sqrt{1-r^2}=\sqrt{1-0^2}=\sqrt{1-0}=1$$

均方根误差以绝对数:磅、美元等等,度量了点沿回归线的散布情况。相关系数从另一方面度量了点相对于 SD 的散布情况,它没有单位。均方根误差通过相关系数与 SD 联系起来。这是我们第三次看到 r 的作用。

- r 描述了相对于 SD,点沿 SD 线的群集程度;
- r 说明了 y 的平均数如何地依赖于 x——x 每增加 1 个 SD,平均来说,y 将只增加 r 个 SD;
- r 通过均方根误差公式,确定了回归预测的精确度。

忠告:如果你超出数据之外作外推预测,或利用回归直线对与研究中的对象范围有所不同的人进行估计,均方根误差将不能告诉你误差有多大。这超出了数学的范围。

习题 B

1. 某法律学校发现了 LSAT 分数与第一年学习成绩间存在下述关系:

LSAT 的平均分数＝35，SD＝5

第一年平均成绩＝65，SD＝10，r＝0.6

招生人员用回归直线从 LSAT 分数预测第一年学习成绩，该直线的均方根误差为下列中的哪一个：

5　10　$\sqrt{1-0.6^2}\times 5$　$\sqrt{1-0.6^2}\times 10$

2.（续上题）

(a)假定你需猜测这些学生之一在第一学年的学习成绩，但却不知道 LSAT 分数，你将用什么方法？

(b)这种方法的均方根误差为

5　10　$\sqrt{1-0.6^2}\times 5$　$\sqrt{1-0.6^2}\times 10$

(c)如果让你知道 LSAT 分数，重做(a)和(b)。

这些习题的答案在第 693 页上。

3. 绘制残差图

统计学家在做回归时常常将残差绘制成图。绘制方法由图 6 指示。散点图中的每一个点都用下述方法转换到第二个图上，它称为残差图。转换时 x 座标不变，y 座标用点的残差——即点在回归线之上(＋)或之下(－)的距离——代替。图 7 表示图 1 的身高与体重散点图的残差图。图 6 和图 7 都表明正残差与负残差互相平衡，用数学语言表示，为：距回归直线的残差平均为 0。残差图还说明：从残差图观察，点没有向上或向下的系统趋势。

残差平均为 0；残差图的回归线是水平的。

其根本理由是所有向上或向下的趋势都已从残差中剔除并表现在回归线中了。

图 7 的残差图没有显示任何格式。作为比较，图 8 给出了一个有明显格式的残差图（数据是虚拟的）。当残差图有这类格式时，使用线性回归很可能是个错误。有时，你可从原散点图直接认定这类非线性。然而，残差图能更灵敏地检测——因为你可以把纵座标放大以进行仔细的观察。残差图在多元回归中是一种非常有用的诊

断手段。例如，从SAT分数和中学GPA预测大学一年级的GPA[②]（多元回归在第12章第3节讨论）。

图6 绘制残差图

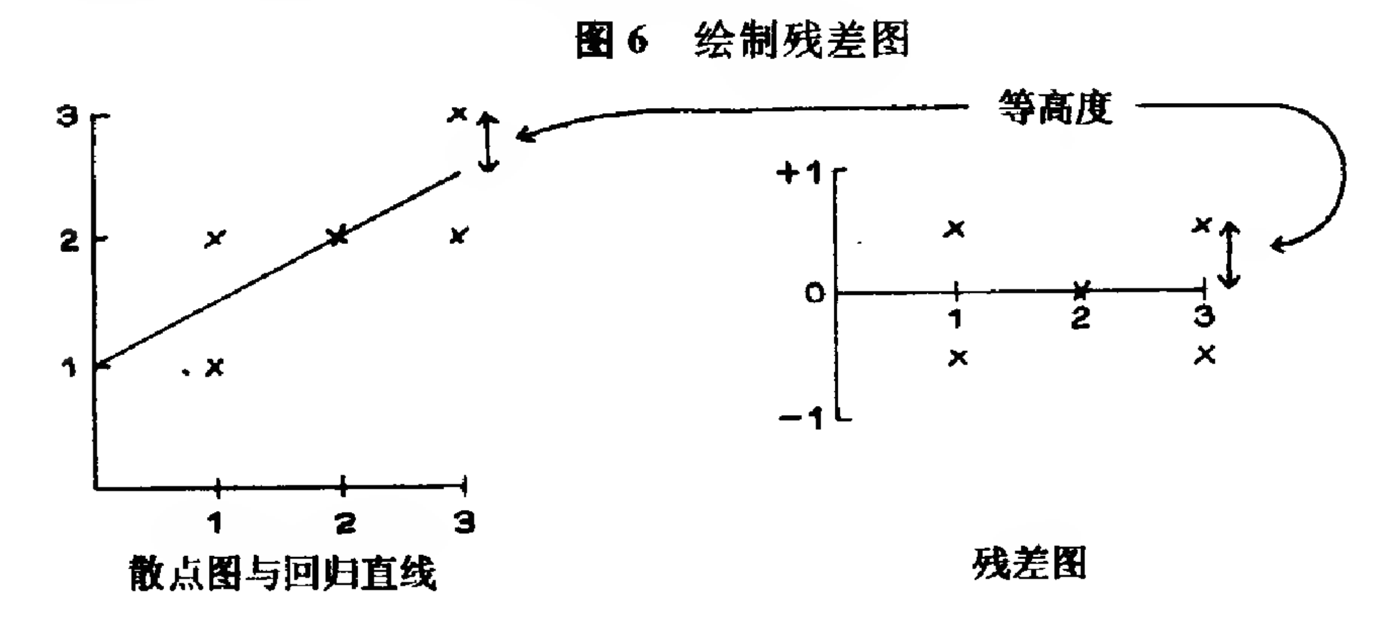


图7 残差图。左边表示HANES样本中988个18—24岁男子的身高和体重的散点图及回归线。残差图在右边，没有趋势和格局。

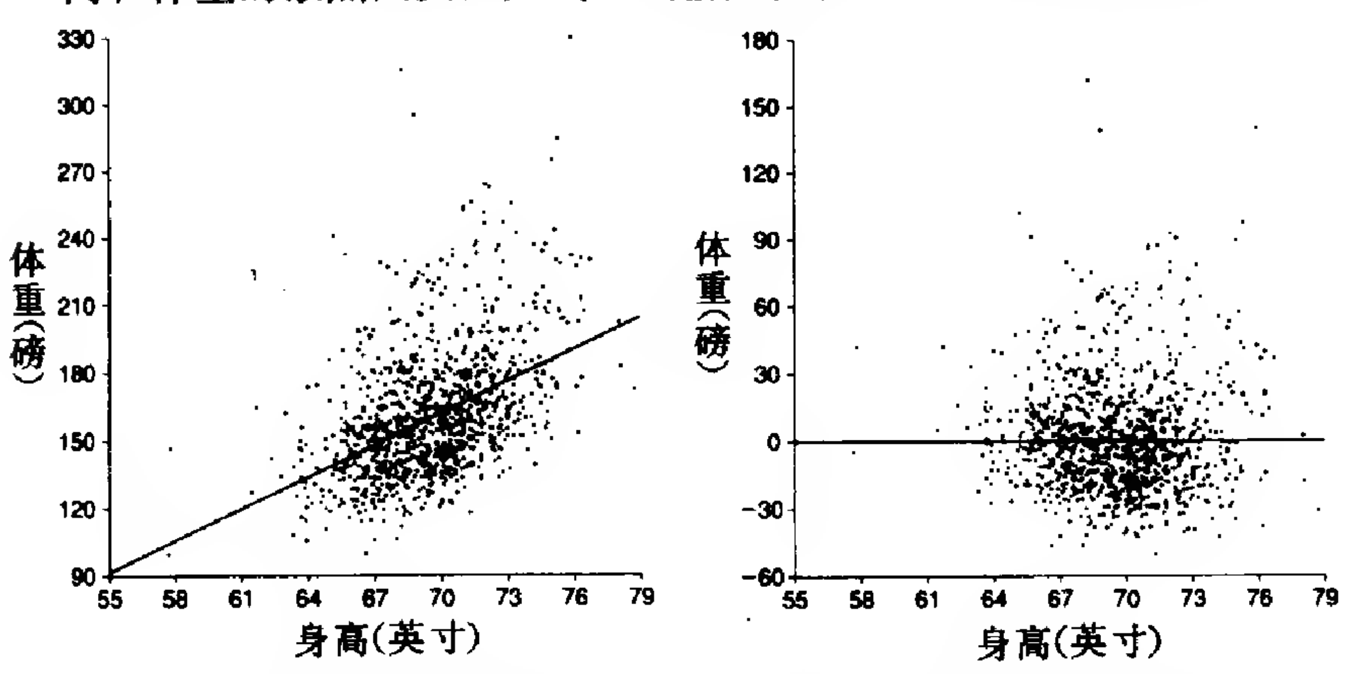
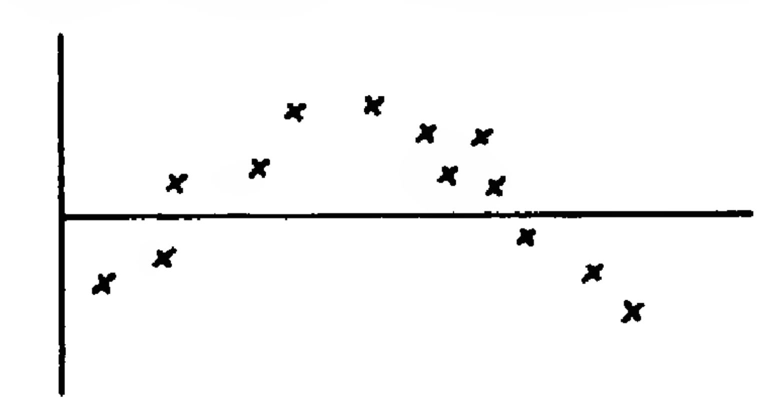


图8 有明显格式的残差图。此时去拟合回归直线是错误的。

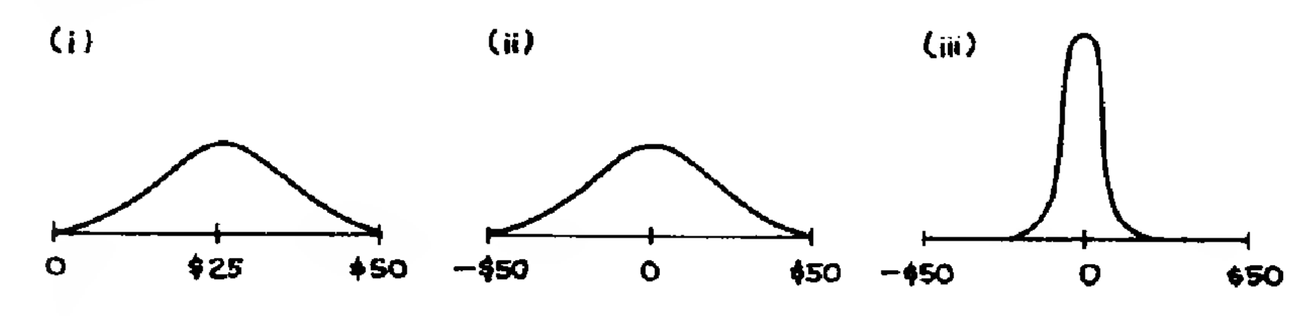


习题 C

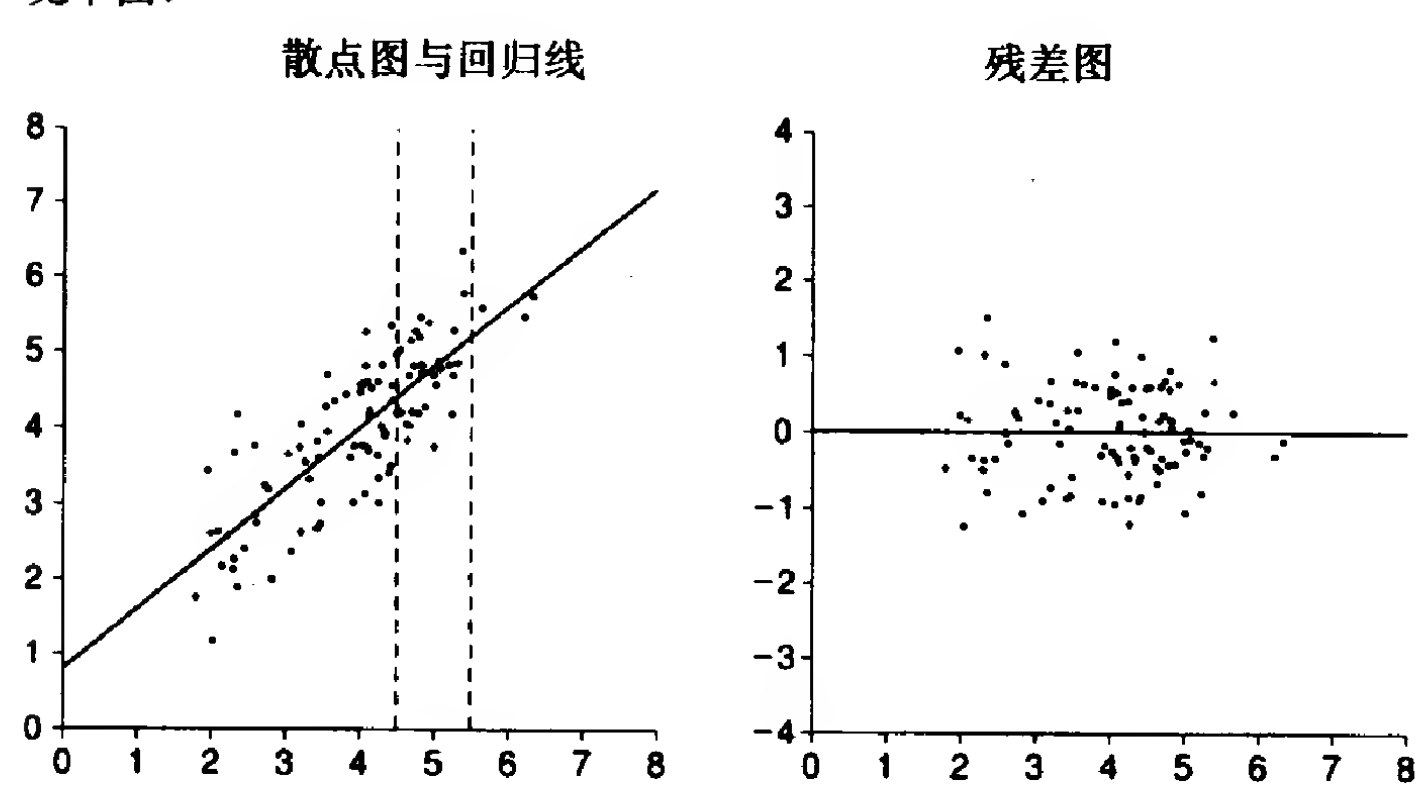
1. 几条不同的回归线被用于预测某种股票的价格(从不同的独立变量来预测),来自各回归线的残差的直方图草绘如下,将描述与直方图相匹配:

(a)均方根误差=5 美元　　(b)均方根误差=15 美元

(c)有某种错误



2. 见下图:



(a)在散点图中,x 和 y 的平均数约为 3、4 还是 5?

(b)x 和 y 的 SD 约为 0.6、1.0 还是 2.0?

(c)相关系数约为 0.4、0.8 还是 0.95?

(d)残差的平均数约为 0、2 还是 4?

(e)残差的 SD 约为 0.6、1.0 还是 2.0?

(f)考虑散点图中 x 座标在 4.5 与 5.5 之间的点,它们的 y 座标的平均数约为 4.6、4.8 还是 5.0?

(g)(f)中点的 y 座标的 SD 约为 0.6、1.0 还是 2.0?

这些习题的答案在第 693 页上。

4. 对纵向条形的考察

图 9 再现了 Pearson 关于 1 078 对父亲和儿子身高的研究的散点图(第 8 章第 1 节)。父亲身高为 64 英寸的家庭在图中用实的纵向条形表示(近似到英寸),在这个纵向条形中的儿子身高的直方图表示在图的下部(实线)。父亲为 72 英寸高的家庭在图中用虚的纵向条形表示,在这个纵向条形中的儿子身高的直方图也表示在图的下部(虚线)。虚线直方图比实线直方图更靠右;平均来说,较高的父亲有较高的儿子。然而,两个直方图的形状很相似,且有差不多相同量的散布[3]。

当散点图中所有纵向条具有相似量的散布时,这个散点图被称作是等方差的(homoscedastic)。图 7 中身高-体重的散点图是等方差的。对应于给定身高,体重的范围在图的中间较大,这只是因为那里有较多的人。从图的一端到另一端,对应于给定身高的体重的 SD 的大小非常相近。一般说来,等方差散点图的残差图是椭圆形的。检验等方差性的最好方法就是观察残差图。("Homo"表示"相同"之意,"scedastic"表示"散布"之意,此处用"homoscedasticity"并不太好,但统计学者们坚持用这个词)。

当散点图是等方差时,整个回归直线上任一点的预测误差都相近。图 9 中,从父亲身高预测儿子身高的回归直线的均方根误差为 2.3 英寸。如果父亲身高为 64 英寸,儿子身高的预测值是 67 英寸,这个预测的误差大约为 2.3 英寸左右。如果父亲为 72 英寸高,儿子身高的预测值为 71 英寸,可能相差大约相同的量,2.3 英寸左右[4]。

作为比较,图 10 给出了收入对受教育水平的异方差性(heteroscedastic)散点图(hetero 意指"不同的")。当受教育水平上升时,平均收入上升,收入的散布也上升。当散点图为异方差时,

图 9　1 078 对父亲和儿子身高的散点图。64 英寸高的父亲的家庭用实纵向条形表示，实线直方图用来描述其儿子的身高。72 英寸高父亲的家庭用虚的纵向条形表示，虚线的直方图用来描述其儿子的身高。虚线的直方图在实线的右面一些——总的说来，较高的父亲有较高的儿子。但两个散点图有相似的形状以及它们的 SD 也几乎相等。

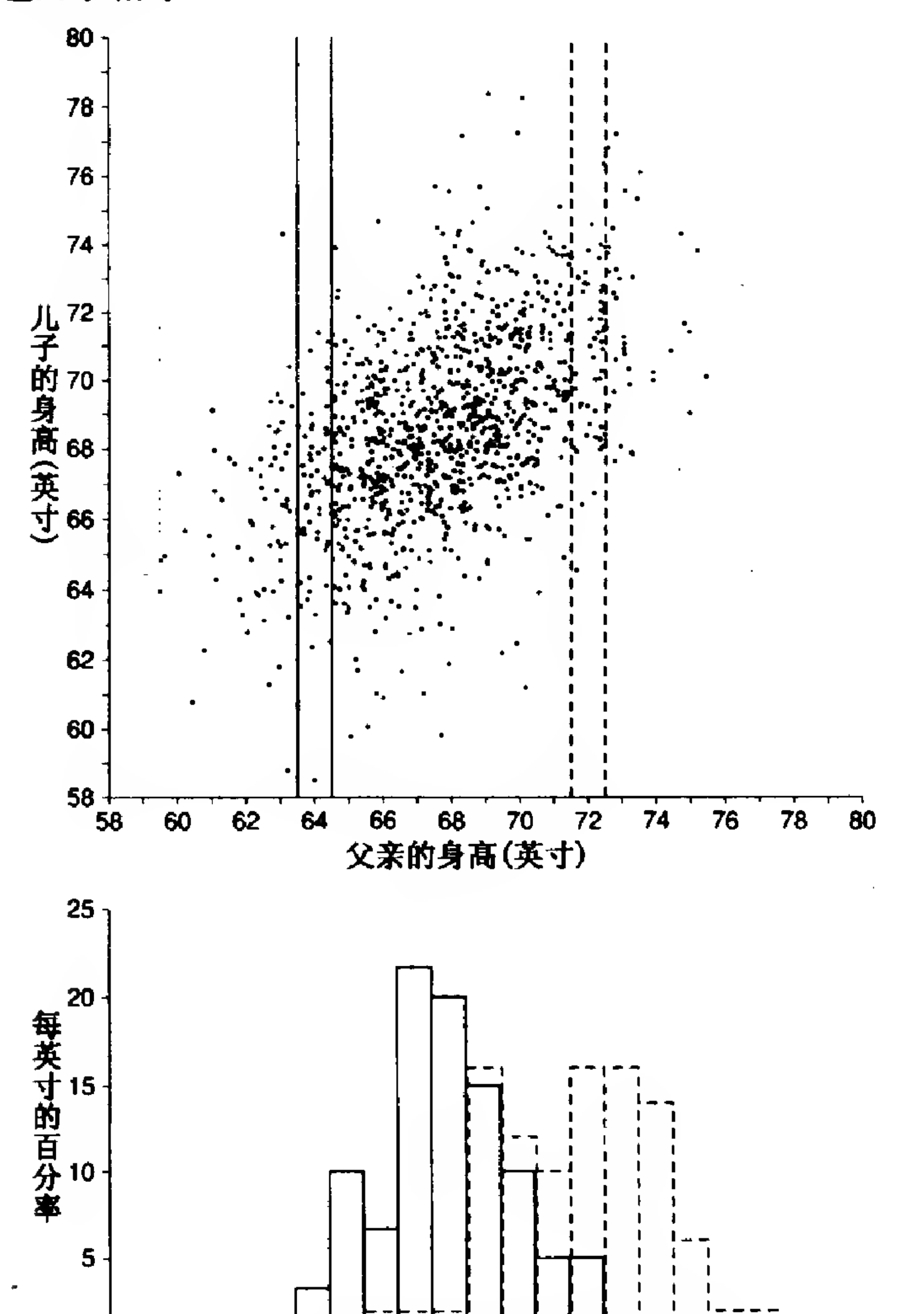
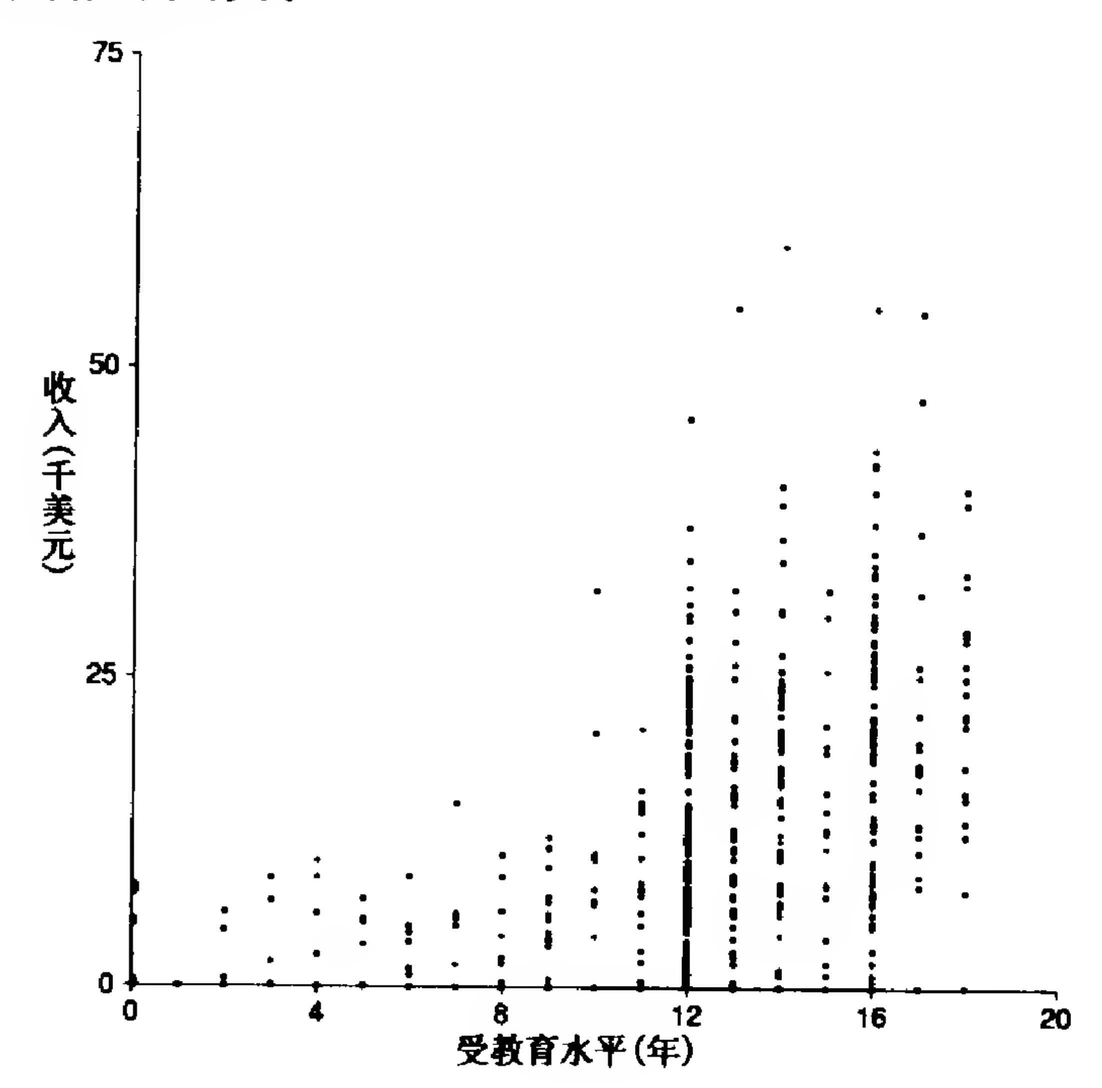


图 10 异方差的散点图。该图表示 1988 年 676 个 25—29 岁加利福尼亚妇女的代表性样本的收入与受教育水平(所完成正规学校教育年数)间的关系[5]。



在散点图的不同地方回归方法偏离不同的量。例如，图 10 中回归线的均方根误差约为 9 600 美元，然而，准确预测受过较高教育的妇女的收入就比较困难。当完成正规学校教育 8 年时，预测误差仅为约 3 400 美元；12 年时，误差上升到 9 100 美元左右；16 年时，误差约为 12 300 美元。在这种情况，回归线的均方根误差只是给出了一些平均误差——对于所有不同的 x 值的。

> 假定散点图是等方差的，且残差中没有一定格式，取一窄纵向条形中的点，它们的 y 值与回归线的偏差在大小上与均方根误差相近。

习题 D

1. 1937 年，Stanford－Binet 智商测验用两种方式(L 和 M)被重新标准化。大量实验对象参加了两种方式的测试，结果概括为：

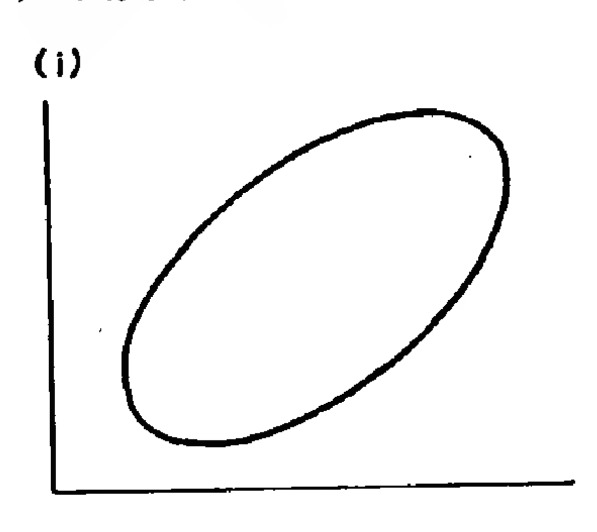
   L 方式平均数≈100，SD≈15

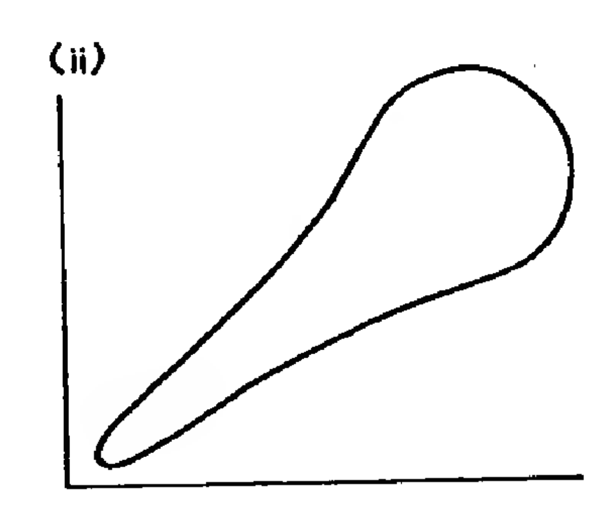
   M 方式平均数≈100，SD≈15，r≈0.80

   (a)正确还是错误并解释之：从 L 方式得分预测 M 方式得分的回归直线的均方根误差约为 9 分。

   (b)假定散点图形如下面的(i)，如果某人的 L 方式得分为 130，回归方法预测其 M 方式得分为 124，正确还是错误并解释之：这个预测可能相差约 9 分左右。

   (c)如果散点图形如(ii)，重复(b)





2. 图 9 的数据可概括如下：

   父亲平均身高≈68 英寸，SD≈2.7 英寸

   儿子平均身高≈69 英寸，SD≈2.7 英寸，r≈0.5

   (a)求出从父亲身高预测儿子身高的回归直线的均方根误差。

   (b)若父亲身高为 72 英寸，预测儿子的身高。

   (c)该预测的可能偏离____英寸左右，或问这个能从给定的信息得出来吗？

   (d)若父亲身高为 66 英寸，重做(b)和(c)。

3. 图 10 的数据可概括如下：

   平均受教育水平≈12 年，SD≈3.5 年

   平均收入≈11 600 美元，SD≈10 500 美元，r≈0.4

   (a)求出从受教育水平预测收入的回归直线的均方根误差。

   (b)预测受教育水平为 14 年的妇女的收入。

(c)该预测的可能偏离____美元左右，或问这能从给定信息得出来吗？

(d)对受教育水平为10年的妇女，重做(b)和(c)。

4. 在一项同卵男性双胞胎的研究中，平均身高约为68英寸，SD约为3英寸。双胞胎的身高间的相关系数约为0.95，散点图为橄榄球状。

(a)未知任何进一步的信息，而你却需猜测这些双胞胎之一的身高，你用什么方法？

(b)求出(a)中方法的均方根误差。

(c)某对双胞胎的其中一位站在你面前，你需猜测这对双胞胎的另一个的身高，你用什么方法？譬如，假定你见到的那位身高为6英尺6英寸高。)

(d)求出(c)中所用方法的均方根误差。

这些习题的答案在693—694页上。

5. 纵向条中正态曲线的应用

在对纵向条形进行研究时，常常可以用正态近似。要使这种应用合理，散点图必须是橄榄球状，散点必须厚厚密集于图的中部而向图的边缘逐渐稀疏。图9就是一个很好的例子。另一方面，如果散点图是异方差的(图10)，或显示了非线性格式，不要使用本节的方法。对于图7中的身高一体重的数据，正态曲线的效果也不太好：点云图不呈橄榄球状；它在上方伸展开来而在底部紧挤。

例1：某法律学校发现在LSAT分数与第一学年成绩之间存在下述关系(对完成了第一学年的学生)：

平均LSAT分数=32，SD=6

第一学年平均成绩=68，SD=10，r=0.60

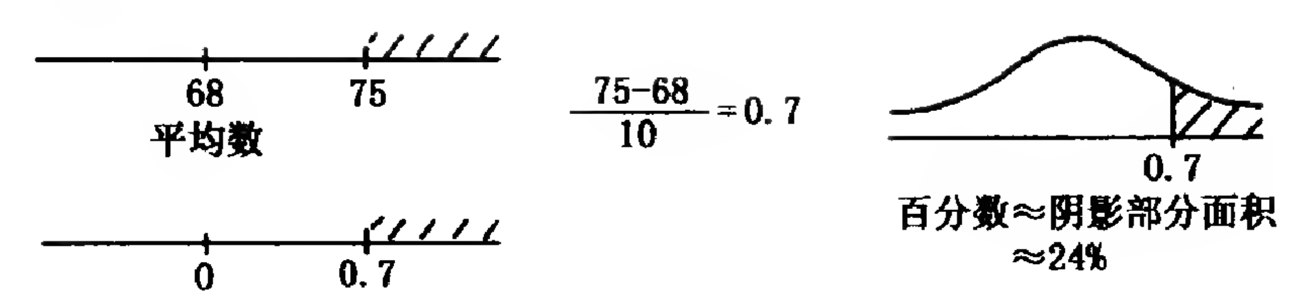
散点图呈橄榄球状。

(a)第一学年分数在75之上的学生约有多少百分数？

(b)在LSAT分数约为35的学生中，第一学年分数在75以上的学生约有多少百分数？

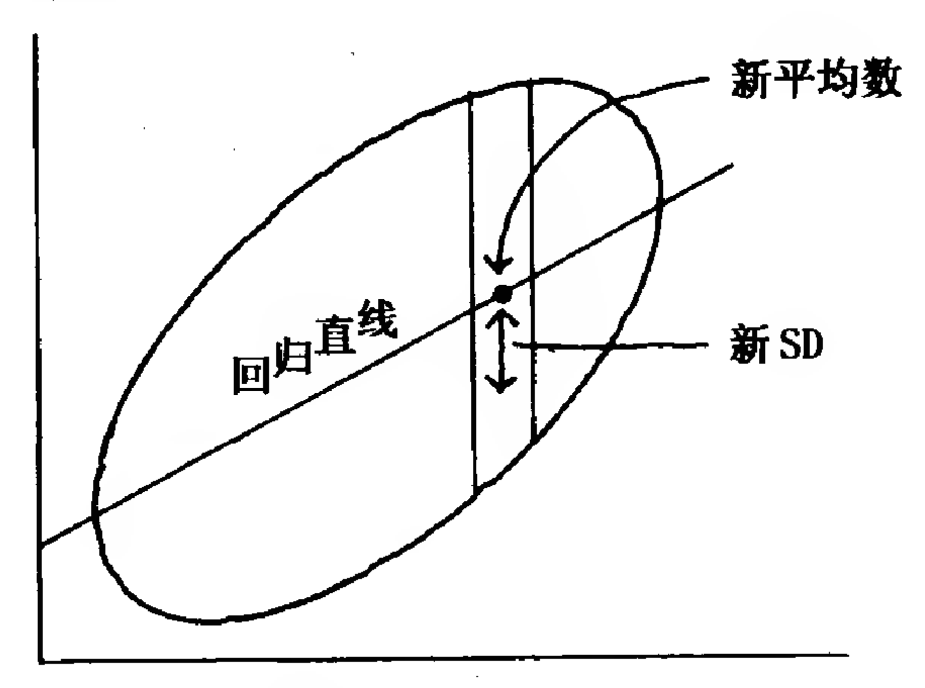
解(a)这是一个直接应用正态近似的问题，LSAT分数和r都

用不到。



(b)这是一个新问题,是关于一群特殊的学生——LSAT 分数为 35 的学生。在散点图(图 11)中,这些学生在一纵向条形中。他们的第一学年分数是一个全新的数据集。为了进行正态近似,需先求出这个新数据集的平均数和 SD。

图 11 橄榄球状散点图。考虑窄纵向条形中的点。它们的 y 值是一个全新的数据集。新平均数由回归方法求得,新 SD 由回归直线的均方根误差得到。纵向条形中具有代表性的 y 值在新平均数附近——加减新 SD。



新平均数。LSAT 分数为 35 的学生的水平比平均数高些。作为一个群体,他们在第一学年考试中的分数要比平均分数高——尽管有相当量的散布,如散点图所示。他们在第一学年考试中的平均分数可用回归方法估计:35 在平均数之上 0.5SD,因此这些学生在第一学年考试中将得到比平均分数高的成绩,约为 $r\times 0.5=0.6\times 0.5=0.3$SD,即为 $0.3\times 10=3$ 分(平均分之上)。因而新平均数为 $68+3=71$。

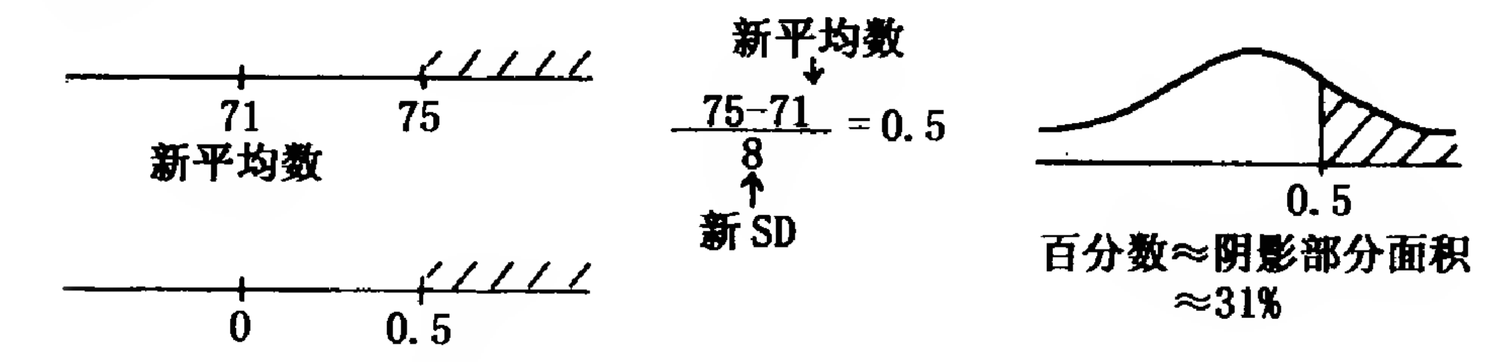
新SD。LSAT 分数为 35 的学生是一个较小而性质更相近的群体，因此他们第一学年分数的 SD 小于 10 分。从图 11 可看出：条状图中的纵向散布程度小于整个图中的。小多少呢？因为这是个橄榄球状图形，围绕回归线的散布程度在每一纵向条形中都大致相同，并且由回归线的均方根误差给定。新 SD 为：

$$\sqrt{1-r^2}\times y\text{ 的 SD}=\sqrt{1-0.6^2}\times 10=8\text{ 分}$$

我们是在从 LSAT 分数预测第一学年成绩，因此这个误差是关于第一学年成绩的：故公式中取 10，而不是 6。

在图中，LSAT 分数沿水平轴标绘，第一学年分数沿纵轴标绘。对整个图纵向 SD 是 10 个第一学年分，但在该纵向条形中仅为 8 分。LSAT 分数约为 35 的代表性学生将有大约 71 的第一学年的分数，误差为 8 分左右。新平均数是 71，新 SD 等于 8。

正态近似是最后一步。这可象通常一样进行，只不过要用新平均数和新 SD。



考虑一个橄榄球状散点图中一个窄纵向条形图中的所有点。它们的 y 值是一个新数据集。其新平均数用回归方法估计，新 SD 约等于回归直线的均方根误差。正态近似可象通常那样进行，但要基于新平均数和新 SD。

习题 E

1. Pearson 和 Lee 取得了约 1 000 个家庭的下述数据：

丈夫平均身高≈68 英寸，SD≈2.7 英寸

妻子平均身高≈63 英寸，SD≈2.5 英寸，r≈0.25

(a)身高超过 5 英尺 8 英寸的妇女有多少百分数？

(b)在嫁给身高为 6 英尺的男子的妇女中，高于 5 英尺 8 英寸的有多少百分数？

2. 同一研究中：

父亲平均身高≈68 英寸，SD≈2.7 英寸

儿子平均身高≈69 英寸，SD≈2.7 英寸，r≈0.50

(a)求高于 6 英尺的儿子有多少百分数？

(b)在身高为 6 英尺的父亲中，其儿子身高超过 6 英尺的有多少百分数？

3. 同一研究中：

男子平均身高≈68 英寸，SD≈2.7 英寸

前臂平均长度≈18 英寸，SD≈1 英寸，r=0.80

(a)求前臂为 18 英寸(近似到英寸)长的男子的百分数？

(b)在身高为 68 英寸的男子中，前臂为 18 英寸长的有多少百分数(近似到英寸)？

这些习题的答案在第 694 页上。

技术性注：如何处理非线性或异方差数据？数据变换常常有所帮助——例如，取对数。图 12 中的左图是 Secchi 深度(水清洁度的一种度量)对叶绿素总浓度(水中水藻的一种度量)的散点图[8]。数据是非线性和异方差的。右图是对同一数据取对数之后：图形更象橄榄球。

6. 复习题

复习题可能包含前面各章的内容

1. 从 x 预测 y 的回归直线的均方根误差是____。

(i)y 的 SD　　(ii)x 的 SD

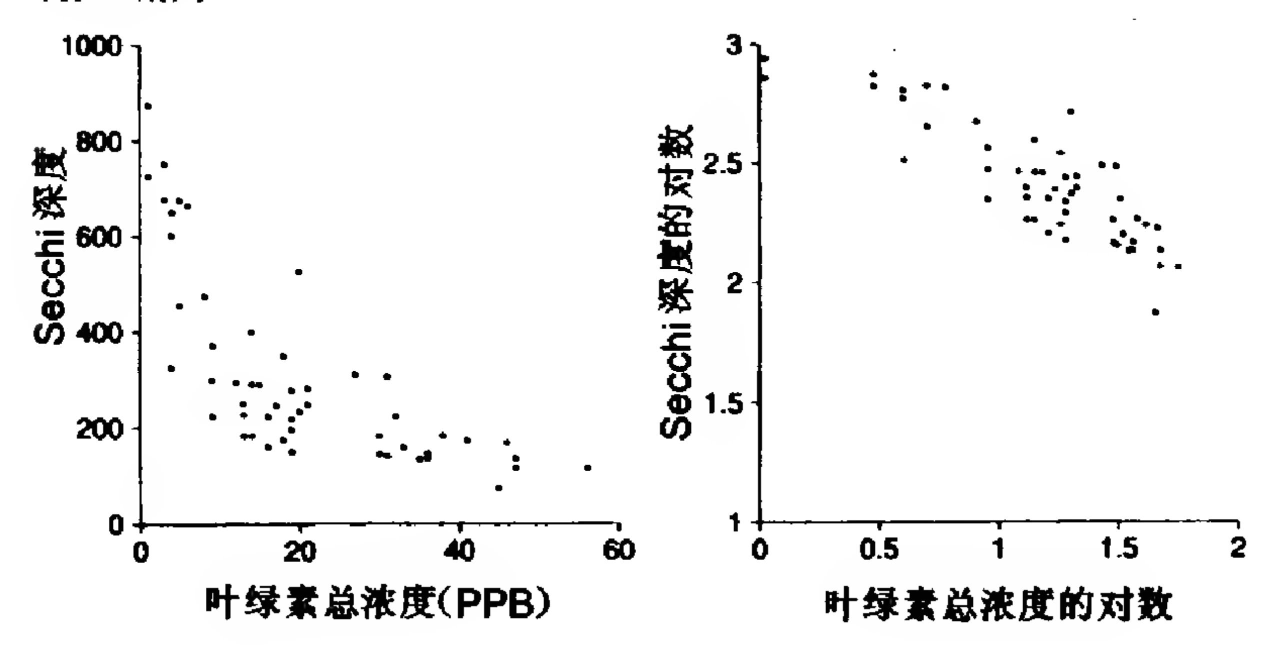
(iii)r×y 的 SD　　(iv)r×x 的 SD

(v)$\sqrt{1-r^2}$×y 的 SD　　(vi)$\sqrt{1-r^2}$×x 的 SD

2. 几条回归直线被用于从不同的独立变量预测在某一公司的年工资，来自各回归线的残差图如下页所示。将描述与残差图相匹配，并解释之。

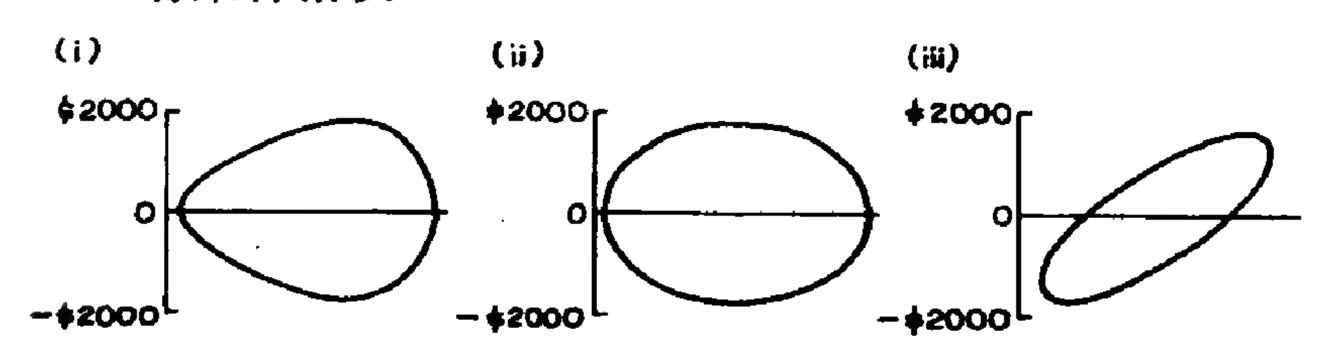
(a)均方根误差=1 000 美元

图 12 左图是 Secchi 深度对叶绿素总浓度的散点图;叶绿素浓度的单位是十亿单位水分之一(ppb)。右图是经对数变换后的数据对数(底为 10)。



(b)异方差的

(c)有某种错误



3. 开发一计算机程序用以从中学的 GPA 预测学院新生的大学 GPA。此程序在一个已知大学 GPA 的班级试用,均方根误差为 3.12。是否有错误? 回答是或否,并解释之。
4. Tuddenham 和 Snyder 得到了 66 个年龄为 6 岁和 18 岁的加利福利亚男孩的下述数据(散点图呈橄榄球状)⑦:
   6 岁时平均身高≈3 英尺 10 英寸,SD≈1.7 英寸
   18 岁时平均身高≈5 英尺 10 英寸,SD≈2.5 英寸,r≈0.80
   (a)对从 6 岁身高预测 18 岁时身高的回归线求均方根误差。
   (b)对从 18 岁时身高预测 6 岁身高的回归线求均方根误差。
5. 在某班级,期中考试平均分数为 50,SD 为 25。期末考试平均分

数为 55,SD 为 15。在期中考试和期末考试分数间的相关系数为 0.60。散点图呈橄榄球状。对每一名学生都用回归方法从期中考试分数预测期末考试成绩。

(a)对约 1/3 的学生来说,期末考试分数的预测值偏离约大于____分。

(b)某学生期中考试分数为 70,预测其期末考试成绩。

(c)这个预测可能偏离____分左右。

从下列备选答案中选择填(a)和(b)中的空。

25　15　12　9　6

6. 对一大课的期中和期末考试成绩的统计分析得到如下结果:

期中考试平均分数≈60,SD≈15

期末考试平均分数≈65,SD≈20,r≈0.50

其散点图呈橄榄球状。

(a)期末考试分数在 80 以上的学生大约多少百分数?

(b)在期中考试分数为 80 的学生中,期末考试分数在 80 以上的约有多少百分数?

7. 某规模较大的大学要求新生参加一组能力测试。数学成绩较好的学生的物理分数也趋于较高。在这两门测试中,平均分数都是 60,SD 为 20(散点图呈橄榄球状)。在数学分数为约 80 分的学生中:

(i)物理分数在 80 以上的正好约为一半

(ii)物理分数在 80 以上的多于半数;

(iii)物理分数在 80 以上的少于半数。

选择一项并解释。

8. 在一项关于中学生的研究中,发现每周做家庭作业的小时数与标准化成绩测试的分数间存在正相关关系。调研者认为做家庭作业能帮助学生准备这些考试。这个结论是由数据得来的吗?回答是或否并简短地解释之。

9. 高空病是由于气压的剧烈变化引起的,它导致在血液中形成氮

气泡，症状表现为剧烈的疼痛，有时出现麻痹导致死亡。在第二次世界大战中，就有飞行员在某些战斗演习中得高空病。高空病发生的条件可以在气压舱中进行模拟。作为一个结果，在训练之初，就让受训的飞行员在这种条件下检验一次。如果他们得了高空病（只是轻微症状），将被从训练中除名，理由是在战斗中他们更可能发作高空病。这种方法被统计学家 J. Berkson 强烈批评，他要求空军重复这种测试——即，对每一个受训的飞行员多重复几次。

(a)为什么 Berkson 会提出这种建议？

(b)举出另一个说明重复是有利的例子。

10. 每年，主要的棒球俱乐部都要授予在第一年比赛中表现突出的选手"本年度新的棒球好手"的称号。从 1949 年到 1987 年，这些年度新的棒球好手的击球平均得分数为 0.285，远高于这些主要俱乐部的击球平均得分 0.257。然而，年度新的棒球好手们在第二年的表现却没有这么好：他们在第二赛季的击球平均得分仅为 0.272。棒球评论家把这称作"第二年滑波"他们的解释是明星运动员被外界活动如产品宣传和在电视中露面分散了精力。你同意还是不同意这种观点？简短地说明之[⑧]。

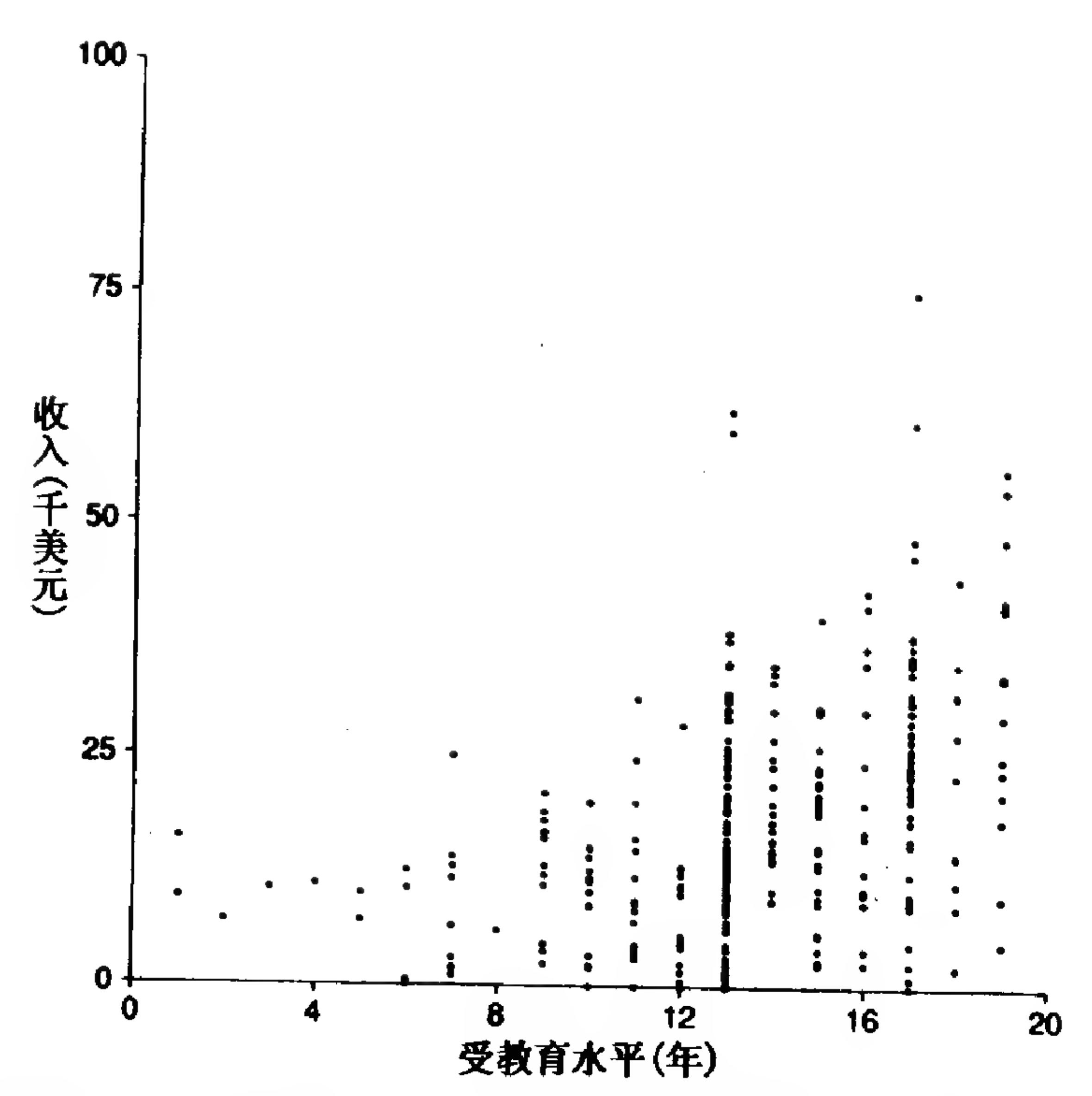
11. 下图是收入对受教育水平的散点图，代表性样本是 1988 年德克萨斯州 25－29 岁男子。有什么错误吗？简短地说明之。（"受教育水平"指所完成的正规学校教育年数，不包括幼儿园。）（图见第 224 页）

7. 小结

1. 当用回归线从 x 预测 y 时，实际值与预测值之间的差异叫残差，或预测误差。

2. 在散点图中，点在回归线之上或之下的纵向距离是回归方法造成的预测误差的图形表示。

3. 回归直线的均方根误差是残差的均方根。它度量回归预测



的精确度。预测偏离一个大小与均方根误差相近的量。对许多散点图，约 68％的预测将正确到一个均方根误差之内；约 95％将正确到两个均方根误差之内。

4. y 对 x 的回归直线的均方根误差可按下式计算：

$$\sqrt{1-r^2}\times y \text{ 的 SD}$$

5. 实施回归后，应绘出残差。如果残差图显示出了一定格式，回归可能是不恰当的。

6. 当散点图的所有纵向条形都有相似的散布量时，该图是等方差的，所有沿回归直线的预测误差在大小上相近。当散点图是异方差时，散点图中不同地方的预测误差是不同的。

7. 考虑橄榄球状散点图中窄纵向条形内的所有点，它们的 y 值是一个新的数据集。新平均数用回归方法估计，新 SD 约等于回归直线的均方根误差。此时基于新平均数和新 SD 的正态近似可象通常那样进行。

12 回归直线

利用某种易受到或大或小误差影响的观察来估计一个量可以并非不恰当地比作某种机会游戏，在这游戏中只能输从不赢，且每一个可能的误差都对应某一损失……然而，那一个特定的损失应归那一个特定的误差本身一点也不清楚。事实上，这个损失的决定至少部分地依赖于我们的判断……在可能函数的无限簇中，最简单的一个似乎占有优势且它无疑应为平方……Laplace 以相同的方式处理这个问题，但他把误差的大小选择为损失的度量。然而，除非我们出差错，他的选择比我们的肯定不会具有更少的随意性。

——C. F. Gauss（德国，1777—1855）[①]

1. 斜率和截距

受教育能带来收入吗？图 1 显示了收入与受教育水平间的关

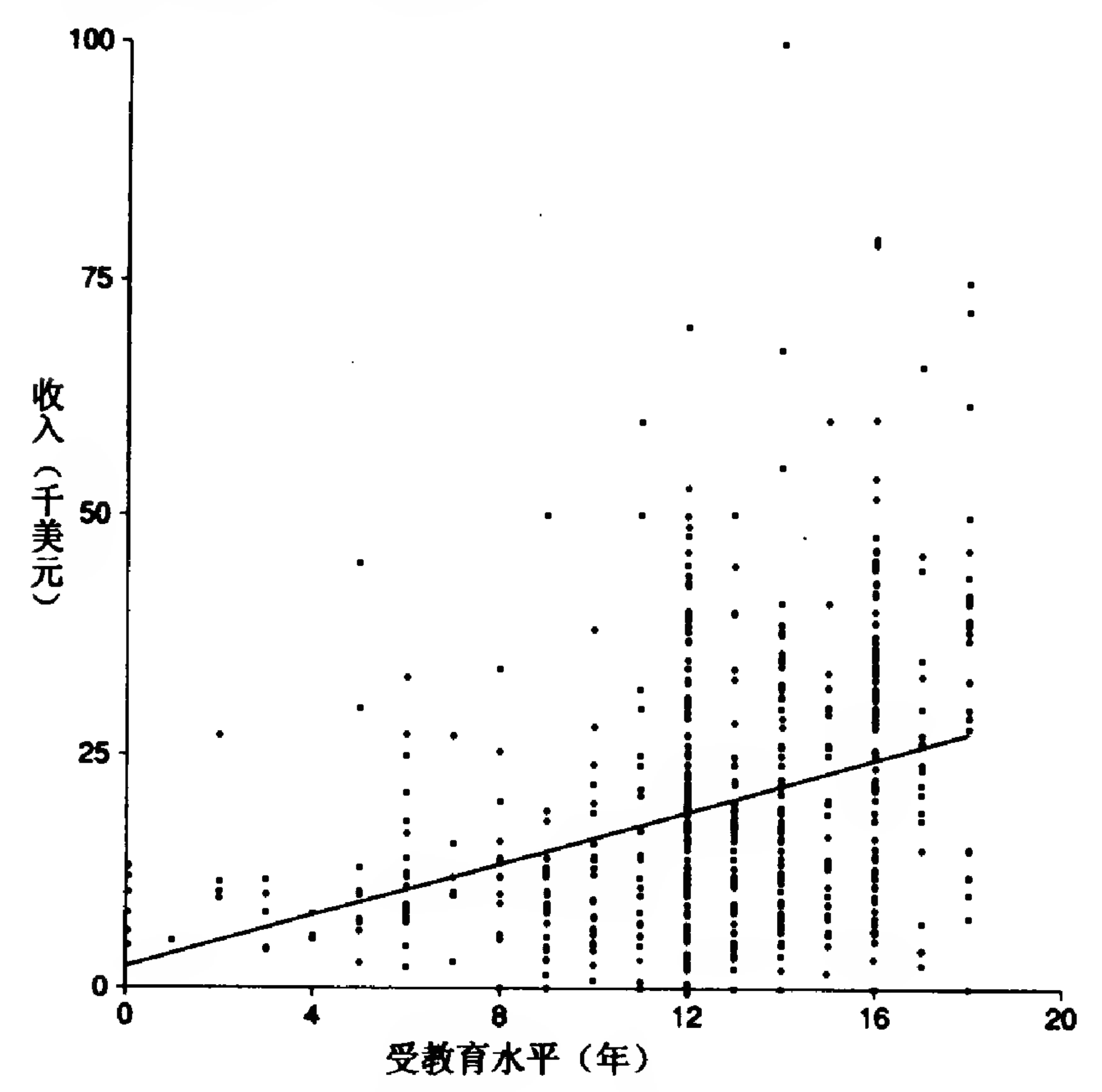
系，代表性样本是1988年637个25—29岁加利福尼亚男子。概括统计量为[②]：

平均受教育水平≈12.5年，SD≈4年

平均收入≈19 700美元，SD≈16 000美元，r≈0.35

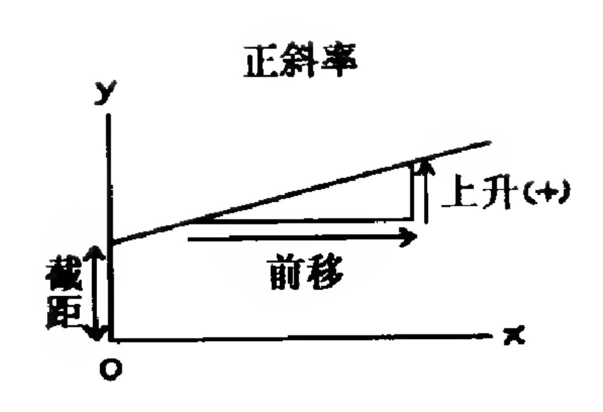
图中的回归直线表示每一受教育水平的平均收入的回归估计。该直线是斜向上的，表明平均说来，收入随受教育水平的提高而增加。

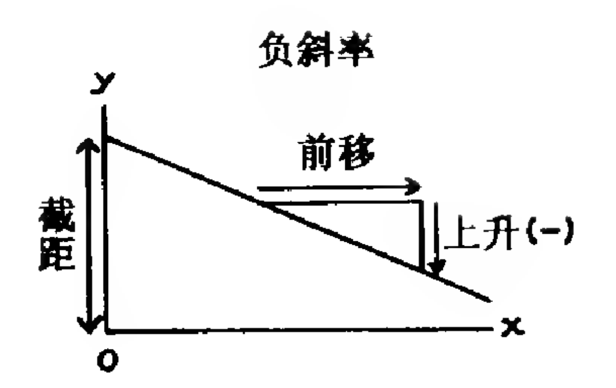
图1 回归直线。散点图表示1988年637个加利福尼亚男子的代表性样本的收入与受教育水平。



任何直线都可用其斜率和截距来描述（见第七章）。y截距是x为0时直线的高度，斜率是x每增加一个单位y增加的比率。斜率和截距的图示见图2。

图 2　斜率和截距





$$斜率=\frac{上升}{前移}$$

对回归线来说，斜率和截距意味着什么?继续以收入—受教育水平为例：与受教育水平每增加 1 个 SD 相应，收入将增加 r 个 SD。即，平均说来，受教育水平增加 4 年，相当于收入增加 0.35×16 000 美元=5 600 美元。因此，每增加受教育一年相当于 5 600 美元/4=1 400 美元。回归直线的斜率是每年 1 400 美元。至此，我们可以看出受教育状况以每年 1 400 美元的比率为他们挣得了收入。

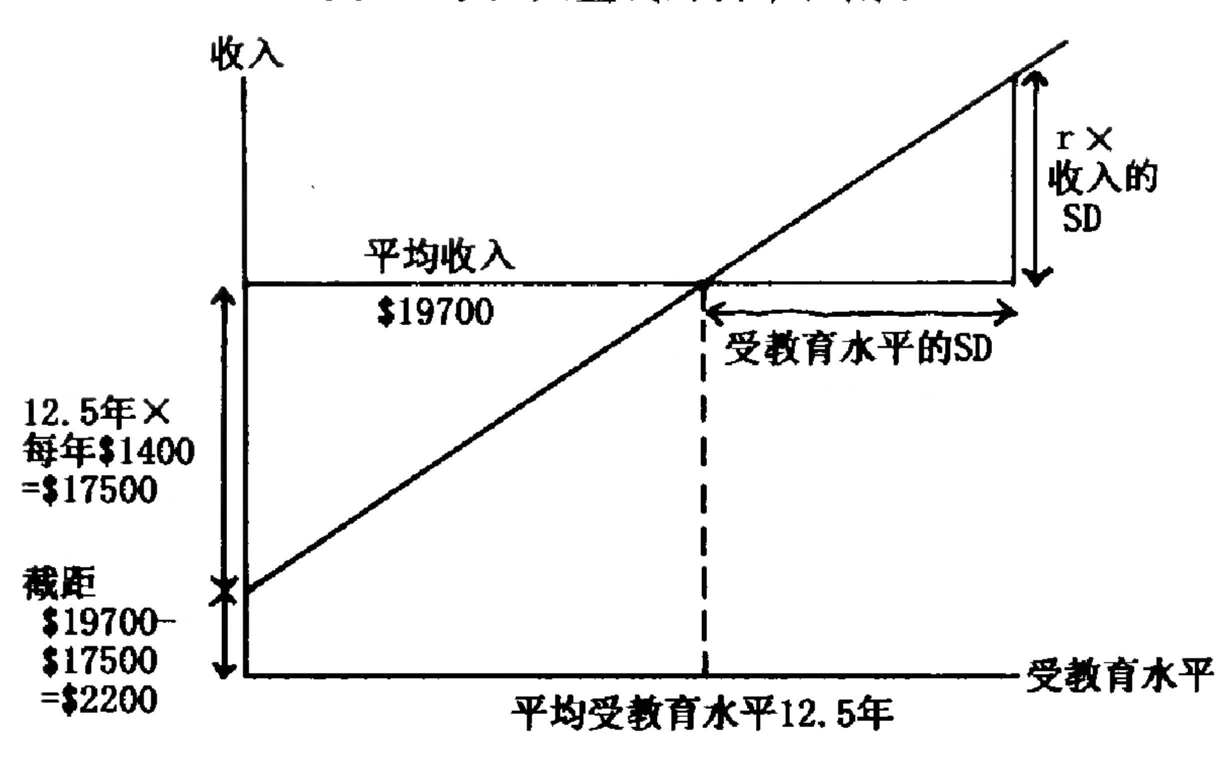
回归直线的截距是 x=0 时回归直线的高，它与受教育水平为 0 年的人相对应。这类人所受教育比平均水平低 12.5 年。因为按照斜率的意义每受一年教育值 1 400 美元，因此可预测未受过任何教育的人的收入比平均水平低约：

12.5 年×1 400 美元/年=17 500 美元

所以，其收入的预测值为 19 700 美元－17 500 美元=2 200 美元。这就是截距：当 x=0 时 y 的预测值(见图 3)。

受教育零年似乎是一个极端，但确有一些人自称未受过任何教育，而他们的收入从 0 到约 12 000 美元；他们的点在图 1 中的左下角。

图 3　求回归直线的斜率和截距



$$斜率=\frac{r\times 收入的\ SD}{受教育水平的\ SD}$$

x 每增加一个单位，y 相应有一个平均变化值。回归直线的斜率说明这个变化值有多大。斜率的计算公式是： $$\frac{r\times y\ 的\ SD}{x\ 的\ SD}$$ 回归直线的截距就是 x＝0 时 y 的预测值。

任何直线都有一个用它的斜率和截距表示的方程：

y＝斜率×x＋截距

回归直线的方程（很自然地）称为回归方程。图 3 中，回归方程为

收入的预测值＝(1 400 美元/年)×受教育年数＋2 200 美元

回归方程并没有引入什么新东西：它只不过是用回归方法由 x 预测 y 的另一种方式。然而，回归方程常用于社会科学。需要预测多个值的研究人员只需计算一次斜率和截距，然后代入方程，就可很容易地算出所有预测值。此外，斜率和截距也有各自的意义。

例 1. 对 1988 年 676 个年龄在 25—29 岁的加利福尼亚妇女，

其收入与教育的散点图见第215页的图10；数据来自现场人口调查。其关系可概括如下[3]：

平均受教育水平≈12年，SD≈3.5年

平均收入≈11 600美元，SD≈10 500美元，r≈0.4

(a)求出由受教育水平预测收入的回归方程；

(b)用回归方程预测受教育水平为8年、12年、16年的妇女的收入。

解：(a)第一步是求斜率。由于受教育水平每增加1个SD，回归直线上升收入的r个SD，因此

$$斜率=\frac{0.4\times 10\ 500\ 美元}{3.5\ 年}=1\ 200\ 美元/年$$

解释：平均而言受教育水平每增加一年，收入增加1 200美元；每减少一年，收入减少1 200美元。

第二步是求出截距。它等于x=0时回归直线的高。换句话说，它是未受过任何教育的妇女收入的预测值。这些妇女所受教育比平均水平低12年，由斜率计算，其收入比平均水平低

$$12\ 年\times 1\ 200\ 美元/年=14\ 400\ 美元$$

因此，她的收入预测为

$$11\ 600\ 美元-14\ 400\ 美元=-2\ 800\ 美元$$

这个值就是截距：x=0时y的预测值(如果离数据的中心越远，回归直线的可靠性就越低，因此，负的截距并不太奇怪)。回归方程为：

$$收入的预测值=(1\ 200\ 美元/年)\times 受教育年数-2\ 800\ 美元$$

(b)将受教育年数8年代入回归方程，得到

$$(1\ 200\ 美元/年)\times(8\ 年)-2\ 800\ 美元=6\ 800\ 美元$$

将受教育年数12年代入，得

$$(1\ 200\ 美元/年)\times(12\ 年)-2\ 800\ 美元=11\ 600\ 美元$$

将受教育年数16年代入，得：

$$(1\ 200\ 美元/年)\times(16\ 年)-2\ 800\ 美元=16\ 400\ 美元$$

解毕。尽管截距是负值，但预测结果对大多数妇女是比较合理的。

本例中，斜率是 1 200 美元/年。受教育水平每增加一年，相应地收入平均增加 1 200 美元。“相应地”一词听起来似乎讲得有点别扭，问题是：收入的差异是由于受教育水平的差异所导致的呢，还是两者都反映了某个第三因素的共同影响？“相应地”这个词就是用来使统计学家谈论回归时不受这类问题的束缚。

当有人介入并改变了 x 值时，常用斜率预测 y 将如何变化。如果数据来自对照试验，这样做是合理的。然而，在观察研究中，这种推断常常是不可靠的——由于混杂的缘故。以例 1 为例，完成了大学教育（受了 16 年教育）的妇女约比仅完成中学教育（12 年）的妇女要多挣 4 800 美元。

如果政府选送有代表性的一组只受过中学教育的妇女去取得大学学位，斜率表明她们的收入平均将增加 4×1 200 美元＝4 800 美元。然而，例 1 是基于调查数据而不是基于对照试验的。调查中的一组妇女受了 12 年教育，另一组不同的妇女受了 16 年教育。这两组除受教育水平外，还有许多其他因素是不同的——如智力、雄心和家庭背景等。

这些因素的作用与受教育水平的作用是混杂在一起的，都反映在斜率中。送人去学习以获得大学学位可能会增加他们的收入，但不一定正好增加满 4 800 美元。为度量大学学位对收入的影响，兴许有必要做对照试验。（尽管许多调研者会采用一种叫多元回归的技术——进一步的有关内容在第 3 节中。）

在观察研究中，回归直线的斜率和截距都只是描述性统计量，它们说明被观察总体中一个变量的平均数是如何与另一个变量的值相联系的。如果研究者改变了 x 的值，不能依靠该斜率来预测 y 相应的变化。

习题 A

1. 对图 1 中的男子，从受教育水平预测平均收入的回归方程是

收入预测值＝(1 400 美元/年)×受教育年数＋2 200 美元

预测受教育水平如下的男子的收入：

(a)8 年——小学

(b)12 年——中学

(c)16 年——大学

2. 菲律宾的国际稻谷研究所培育了一种杂交稻谷 IR8，带来了热带农业的“绿色革命”。除此之外，他们还就肥料对稻谷产量的影响进行了详细的研究。这些试验涉及到大量的试验小块地(大小约为 20 平方码)。每一小块都种 IR8，并施以由研究人员确定的一定量的氮肥(施肥量从 0 到 1 磅)。收获时，测定了产量并发现它与施用的氮肥量有关。在某次这样的试验中，稻谷产量与氮肥施用量之间的相关系数为 0.95，回归方程为[4]：

稻谷产量预测值＝(20 盎司稻谷/盎司氮肥)×(氮肥用量)＋240 盎司

(a)预计未施肥的那块地可产约____稻谷。

(b)氮肥用量每增加 1 盎司预计可增产稻谷____。

(c)预测施肥量如下时稻谷的产量

3 盎司氮肥　　4 盎司氮肥

(d)这是观察研究还是对照实验？

(e)实际上施用的化肥量只有如下值：0 盎司、4 盎司、8 盎司、12 盎司、16 盎司，你认为对施 3 盎司氮肥时的预测值可信吗——尽管这个特定量从不施用？

(f)你认为施氮肥 100 盎司时的预测值可信吗？

3. 1988 年，35—54 岁具有专职工作男子的收入与受教育水平间的关系可概括如下——以种族对照[5]：

白人：平均受教育水平≈13.5 年，SD≈3 年

平均收入≈33 900 美元，SD≈22 000 美元，r≈0.39

黑人：平均受教育水平≈12.3 年，SD≈3 年

平均收入≈24 200 美元，SD≈16 000 美元，r≈0.42

预测具有高中学历(12 年教育)的白人男子的收入。预测同样学历的黑人的收入。

4. 对 HANES 样本(第 64 页)中 25—34 岁的男子,从受教育水平预测身高的回归方程为[⑥]:

身高预测值=(0.25 英寸/年)×(受教育年数)+66.75 英寸

预测受教育水平分别为 12 年和 16 年的男子的身高。上大学增加了男子的身高吗? 请解释。

5. 求出由父亲身高预测儿子身高的回归方程(概括统计量见第 191 页)。

6. 求出由儿子身高预测父亲身高的回归方程(数据同习题 5)。

这些习题的答案在第 694—695 页上。

2. 最小二乘法

第 10 章从一个角度讨论了回归,本章第一节用回归方程重又探讨了同一话题。本节再从另一观点第三次提出同一论题(对统计学家来说,回归是一种重要的方法)。有时散点图中的点似乎沿一条直线伸展。本节中讨论的问题是如何找到一条直线,它最佳地拟合这些点。通常这涉及某种折衷办法:移动这条直线以靠近某些点将增大它与其他点的距离。要解决这个矛盾,有两个步骤是必不可少的:第一,定义一个从直线到所有点的平均距离;第二,在附近移动这条直线直至这个平均距离尽可能地小。

为使问题更具体,假定这条直线是用作由 x 预测 y 的。于是,每一个点所产生的误差就是从该点到直线的纵向距离(见第 11 章第 1 节)。统计学中,定义平均距离的常用方法是取这些误差的均方根。平均距离的这种度量称为直线的均方根误差。(这首先由 Gauss 提出,他很谨慎地指出这仅仅是一个定义。见本章开头的引语。)

第二个问题是如何在附近移动这条直线以使均方根误差最小,这也由 Gauss 解决:

在所有直线中,使由 x 预测 y 的均方根误差最小的那条直线是回归直线。

由于这个原因,回归直线常常又称为最小二乘直线:将误差平方

以计算均方根误差，回归直线使均方根误差尽可能小。

现在来看一个例子。Robert Hooke（英国，1653—1703）确定了弹簧长度与其负重间的关系。他在弹簧一端悬挂不同重量的重物并观察其变化：当他增加重物的重量时，弹簧拉长；当他减轻重量时，弹簧缩短，且这个关系是线性的。

令 b 为弹簧不挂重物时的长度。然后在弹簧一端悬挂 x 千克的重物，如图 4 所示，这时弹簧将伸长到一个新的长度。根据 Hooke 定律，伸长的长度与重量 x 是成比例的，因此弹簧的新长度是

$$y=mx+b$$

在这个方程中，m 和 b 都是取决于弹簧的常量，它们的值未知且必须通过实验数据来估计。

图 4 Hooke 定律：弹簧伸长的长度与重物重量成比例。

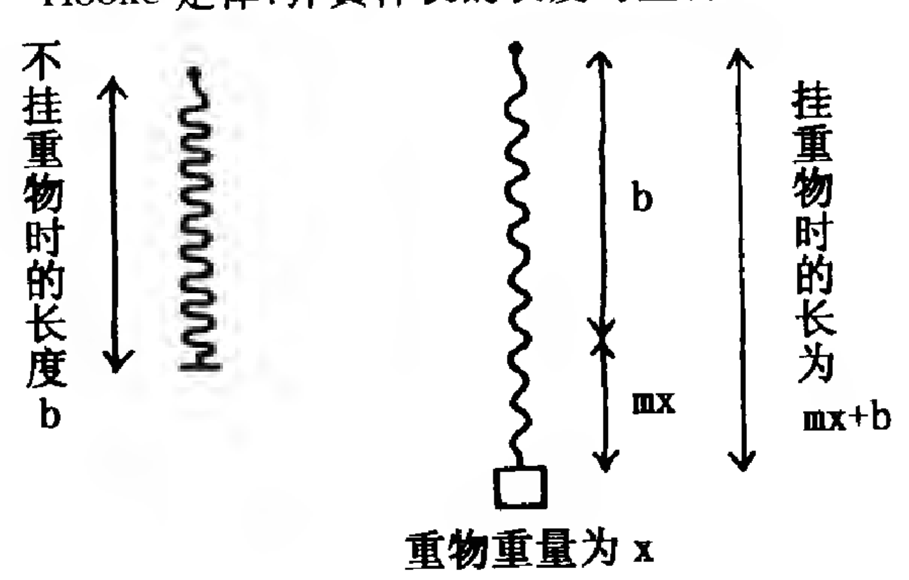


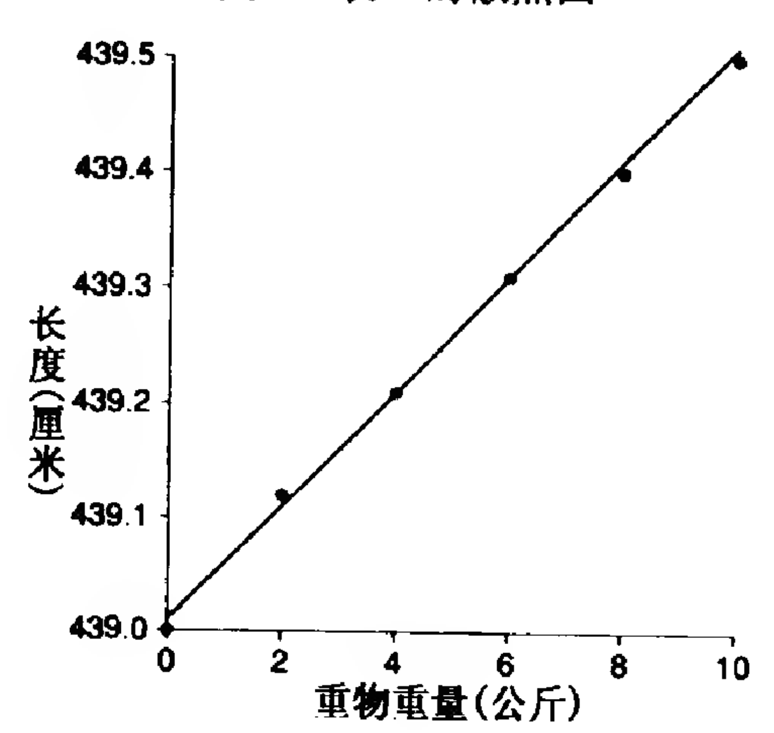
表 1 是关于 Hooke 定律的一次实验的结果，在那次实验中把不同大小的砝码悬挂于一段钢琴弦的末端[7]。第一列表示重量，第二列表示测量长度。当重物为 20 磅时，这个“弹簧”仅伸长约 0.2 英寸（10 千克≈22 磅，0.5 厘米≈0.2 英寸），看来钢琴弦不是很有弹性。

表 1　关于 Hooke 定律的数据

重量(千克)	长度(厘米)
0	439.00
2	439.12
4	439.21
6	439.31
8	439.40
10	439.50

表 1 中数据的相关系数为 0.999，非常接近于 1。因此，这些点几乎成一条直线(图 5)，恰如 Hooke 定律预测的那样。与线性的很小的差异可能是由于测量误差引起的——不论砝码还是长度都没有完全精确地测量。

图 5　表 1 的散点图



我们的目标是估计符合钢琴弦的 Hooke 定律方程中的 m 与 b

$$y = mx + b$$

这个方程的图是一条理想的直线，图 5 的散点图是它的近似图。如果图 5 中的点正好完全落在某条直线上，我们将取这条直线作为理想直线的近似，其斜率是 m 的估计值，截距是 b 的估计值。

麻烦在于这些点没有完全在一条直线上，穿过散点图可以绘

出很多条不同的直线，它们有着稍微不同的斜率和截距。该用哪一条直线呢？Hooke 方程从重量预测长度，如前面所讨论的，很自然地应选择 m 和 b 使预测误差的均方根误差最小：这就是最小二乘法。这样的直线 $y=mx+b$ 就是回归线[8]。换句话说，Hooke 定律中的 m 应估计为回归直线的斜率，b 估计为截距。这叫最小二乘估计，因为它们使均方根误差最小。计算得：

$$m\approx 0.05\text{cm/kg},\ b\approx 439.01\text{cm}$$

最小二乘法估计该弹簧不悬挂重物时的长度是 439.01cm，每挂一千克重量估计弹簧将伸长约 0.05cm。因为这是基于对照实验得出的结果，因此无需对结论作更多的解释。研究人员增加重量，弦将伸长；取下重物，弦将恢复到原来的长度。这个过程可按需要多次重复进行。毫无疑问这里存在着因果关系，用不着“相关”这种语言。

最小二乘法和回归方法包含了同样的数学原理，但来龙去脉可能有所不同。在有的领域，研究人员在估计参数时谈论“最小二乘方”，比如估计 Hooke 定律中的 m 和 b。在另一些领域，研究人员研究两个变量比如收入与教育间的关系时谈论回归。

一个技术问题：弹簧未挂重物时长度的最小二乘估计是 439.01cm，比未挂重物时长度的实际测量值(439.00cm)略大一点。统计学家认为这个最小二乘估计是更可信的。为什么呢？因为最小平方估计是利用了所有六个测量值，而不是一个，得出的。有的测量误差可能互相抵销。当然，这六个测量值是由一个好的理论——Hooke 定律联系在一起的。没有这个定律，最小二乘估计没有多大用处。

3. 回归有意义吗？

在任何散点图中都可以绘出回归直线。然而，有两个问题要提出。首先，变量间存在非线性关系吗？若是，回归直线非常可能将人引入歧途(这在第 10 章第 2 节已讨论过)。即便相关看来是线性

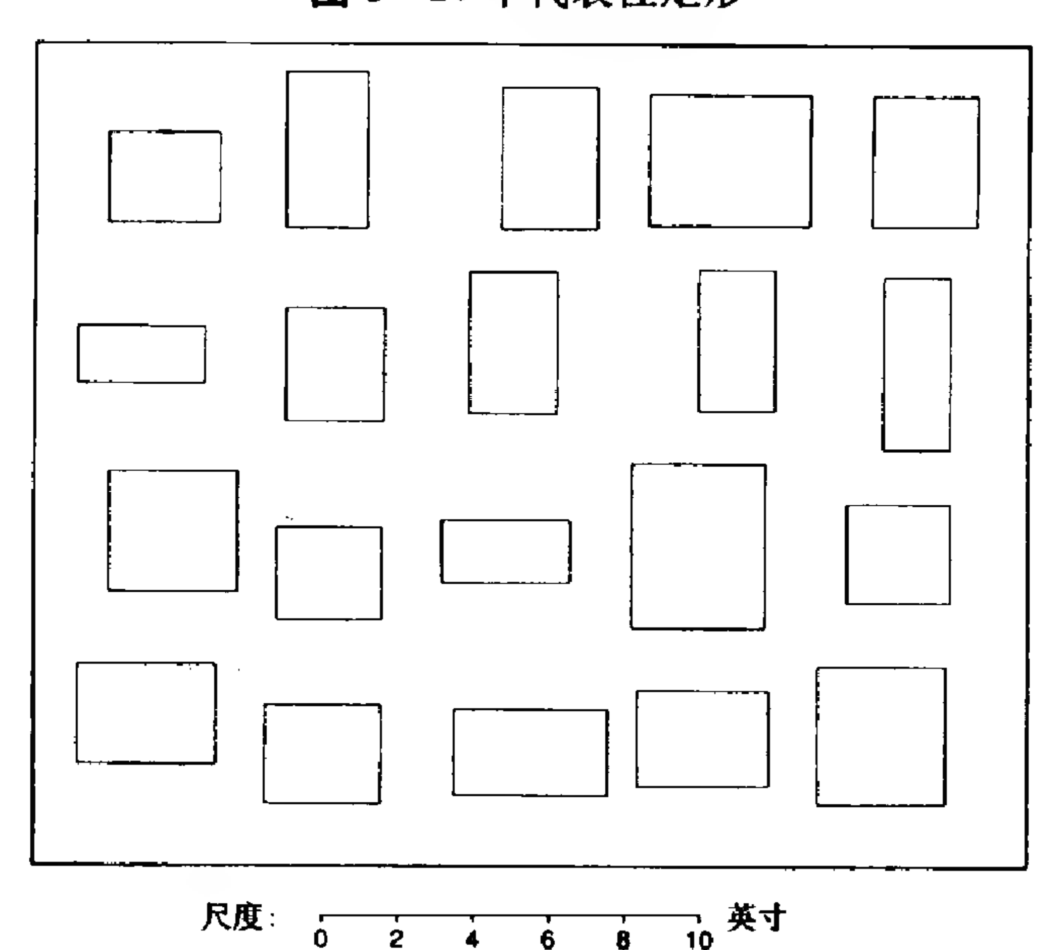
的，还存在第二个问题：回归有意义吗？这个问题比较困难。回答这个问题必须了解产生数据的机制，如果对这一机制没有充分的理解，拟合一条回归直线可能是一场灾难。

举个例子，假定某研究者不知道矩形面积的计算公式。采用经验方法他画了20个代表性矩形，如图6所示。他认为矩形的面积应依赖矩形的周长。对每一矩形他测量了面积和周长。有关结果的散点图如图7所示。相关系数得出为0.98——与Hooke定律中的相关系数几乎一样好。该研究者得出结论他真正地知道了某些东西，他的回归方程为：

面积预测值=(1.60英寸)×周长−10.51平方英寸

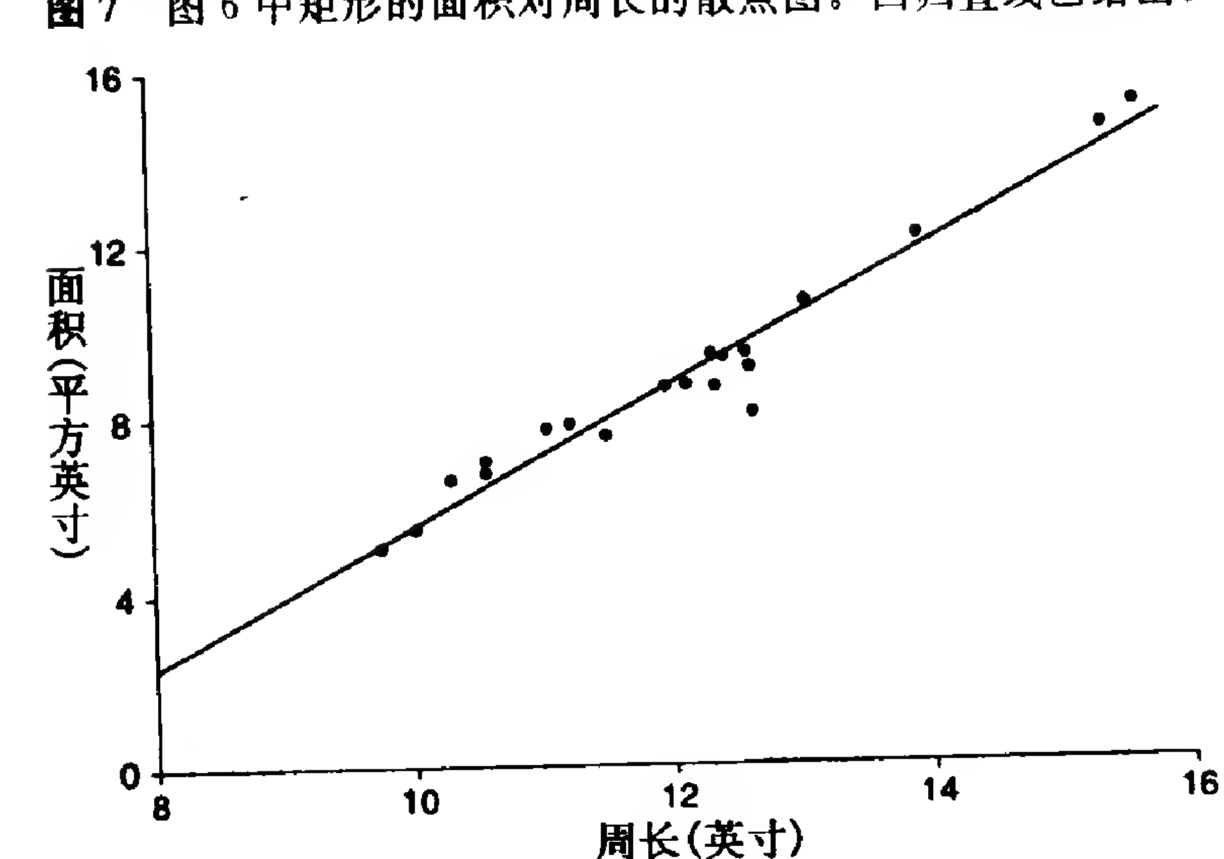
(面积用平方英寸而周长用英寸测量。)回归线如图7所示。

图6 20个代表性矩形



算术运算都是正确的。但这个研究者考虑问题如此粗糙使得方程非常可笑。他应该考虑矩形的长和宽这两个变量，它们确定了矩形的面积和周长：

图 7 图 6 中矩形的面积对周长的散点图。回归直线已给出。



面积＝长×宽，周长＝2(长＋宽)

这个傻瓜的研究者用回归将永远发现不了这个关系。

当然，这仅是一个虚构的例子。但许多研究者在并不真正地知道发生的情况下，就给散点图配回归直线。这就产生了很多问题。在考虑回归时，问一下你自己问题是否更象 Hooke 定律。还是更象面积与周长。

技术性注：第 1 节的例 1 给出了由受教育水平预测收入的回归方程，简要回顾一下，这是描述收入与受教育水平之间关系的一个好办法。但是，1 200 美元/年的直线斜率解释成因果关系可能是不合理的：如果让人去学校再接受一年教育，那样会使他们平均多挣 1 200 美元，产生问题—— 其他变量的作用可能与受教育水平的作用混杂在一起。

许多研究人员用多元回归来控制其他变量。例如，他们可以提出利用各个体的父母社会经济地位的度量，并拟合如下形式的多元回归方程：

$$y = a + b \times E + c \times s$$

其中

y＝收入的预测值，E＝受教育水平

$$s = \text{父母地位的度量}$$

系数 b 可以解释为控制了父母地位的影响后，受教育水平的作用表现。

这可能给出有意义的结果。但这种方法同样也会产生无价值的东西。我们来看正在研究矩形面积的那个虚构的研究者。他也可能用对角线长来度量矩形的形状，通过多元回归控制矩形形状的影响（当然，用对角线长度量形状不是好方法；但也没有人知道如何度量社会经济地位）。该研究者可能拟合形式如下的多元回归方程。

$$\text{预测面积} = a + b \times \text{周长} + c \times \text{对角线长}$$

他认为 b 度量了控制形状的影响后周长的作用。结果，他越做越糊涂。周长和对角线确实能确定面积，但仅仅是通过一个复杂的非线性的公式。多元回归是一种非常有用的技术，但它代替不了对数据间内在关系的了解[9]。

习题 B

1. 对第 2 节中讨论的钢琴弦问题（表 1 和图 5），预测悬挂如下重量（如果可能）时弦的长度：3 千克、7 千克、50 千克。
2. 下表数据为 1960 至 1986 年人均可支配收入和个人消费支出（按 1982 年美元计）。

   (a)求出 r 和由收入预测消费的回归方程。

   (b)绘出残差。

年份	收入	消费
1960	6 036	5 561
1961	6 113	5 579
1962	6 271	5 729
1963	6 378	5 855
1964	6 727	6 099
1965	7 027	6 362
1966	7 280	6 607
1967	7 513	6 730
1968	7 728	7 003
1969	7 891	7 185
1970	8 134	7 275
1971	8 322	7 409

续表

年份	收入	消费
1972	8 562	7 726
1973	9 042	7 972
1974	8 867	7 826
1975	8 944	7 926
1976	9 175	8 272
1977	9 381	8 551
1978	9 735	8 808
1979	9 829	8 904
1980	9 722	8 783
1981	9 769	8 794
1982	9 725	8 818
1983	9 930	9 139
1984	10 419	9 489
1985	10 622	9 830
1986	10 947	10 142

注:表中金额为按 1982 年美元计的人均值。

来源:《总统经济报告》,1988 年,表 B-27。

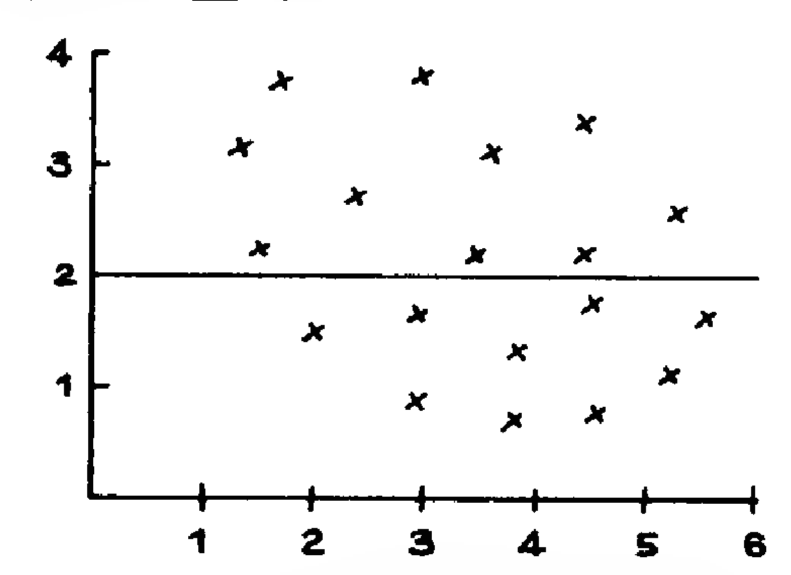
这些习题的答案在第 695 页上。

4. 复习题

复习题可能包含前面各章的内容。

1. (a)下图中直线的均方根误差约为 0.1、0.3 还是 1?

   (b)图中直线是回归直线吗?



2. 利用下面资料,求出由期中考试分数预测期末考试分数的回归方程:

   期中平均分数=70,SD=10

   期末平均分数=55,SD=20,r=0.60

3. 对第 2 节中讨论的钢琴弦(表 1 和图 5),如果可能,预测悬挂如下重量时弦的长度:

(a)1 千克　　(b)5 千克

(c)9 千克　　(d)100 千克

4. 某研究人员建立了一个由车的长度(英寸)预测车的重量(磅)的回归方程,其斜率约为:

每英寸 3 磅　　每英寸 30 磅

每磅 3 英寸　　每平方英寸 30 磅

每英寸 300 磅　　每千克 300 厘米

选择一个答案并解释之。

5. 现有一个由身高预测体重的回归方程:

体重预测值=(4 磅/英寸)×身高−130 磅

其中体重与身高分别以磅和英寸为单位。如果换算为公制(1 英寸≈2.5 厘米,1 磅≈0.45 千克)回归方程是怎么样的?

6. 数据如题 5,另一研究人员想得到从体重预测身高的回归方程,她从上面的回归方程解得:

身高预测值=($\frac{1}{4}$英寸/磅)×(体重+130 磅)

这样做对吗?如果你需要计算的话,有关数据如下:

平均身高≈68 英寸,身高 SD≈2.5 英寸

平均体重≈142 磅,体重 SD≈25 磅,r≈0.40

7. 对 HANES 样本中 25—34 岁的男子,身高与收入间的关系可概括如下[10]:

平均收入≈14 900 美元,SD≈7 200 美元

平均身高≈70 英寸,SD≈3 英寸,r≈0.2

由身高预测收入的回归方程,是什么?对它作说明。

8. 对 HANES 样本中 18—24 岁的男子,由体重预测身高的回归方程为:

身高预测值=(0.047 英寸/磅)×体重+62.4 英寸

(身高以英寸、体重以磅为单位)。若某人体重增加 20 磅，他将长高

20 磅×0.047 英寸/磅≈0.9 英寸吗?

若否，斜率意指什么?

9. 在一项虚构的关于父母收入与子女智商间关系的大型研究中，得到了如下结果:

平均收入≈21 000 美元，SD≈15 000 美元

平均智商≈100，SD≈15，r≈0.50

在分析数据时，按如下方式绘制了一个图:对每一收入组(0—1 000 美元，1 000—2 000 美元，2 000—3 000 美元，等等)，计算出父母收入在该组的子女智商的平均数，并画在该组的中点(500 美元，1500 美元，2500 美元，等等)的上方。发现图中的点非常接近地拟合一条直线，该直线的斜率(每一美元的智商分数)将约为:

2 000　　1 000　　500　　100　　1/100　　1/500　　1/1 000　　1/2 000　　从已知信息不能得出

从中选择一个答案，并简短地解释之。

10. 题 9 的研究中的某一孩子的智商为 110，但其父母收入情况的信息遗失了，题 9 中所绘直线在 41 000 美元处的高度对应于智商分 110。41 000 美元是否是其父母收入的一个好的估计?或它可能太高?太低?请作解释。

11. 摄取食盐会导致高血压吗?关于这个问题的一项大型研究在 32 个国家的 52 个中心进行[11]。每个中心征集了 200 个实验对象，按年龄和性别分为 8 个组，对他们的食盐摄入量、血压、及几个可能的混杂变量进行了测量。在对年龄、性别及混杂变量作出调整后，25 个中心发现舒张压和食盐摄取量间是正相关;27 个中心发现是负相关。这些数据能证明食盐会导致高血压吗?回答是或否，并简短地解释。

12. 某专家证人提供了证词称“回归可以代替对照实验，它能给出

一个变量对另一个变量的影响的精确估计”[12]。简短地进行评述。

5. 小结

1. 回归直线可用两个描述性统计量确定：斜率和截距。

2. y 对 x 的回归直线的斜率，是对应于 x 的每一单位变化，y 所发生的平均变化，它等于 r×(y 的 SD)/(x 的 SD)。

3. 回归直线的截距等于 x 为 0 时 y 的回归估计。

4. y 对 x 的回归直线的方程为

$$y = 斜率 \times x + 截距$$

5. 回归方程通过代入可用来作出所有回归预测。

6. 在所有直线中，回归直线使由 x 预测 y 的均方根误差最小；由于这个原因，它又常被称为最小二乘直线。

7. 有时，两个变量被认为可通过某线性关系相联(例如，Hooke 定律中的长度和重量)。统计问题是估计该直线的斜率和截距。最小二乘估计是回归直线的斜率和截距。

8. 在对来自观察研究的数据配以回归直线时，把斜率解释为 x 的每一单元变化所导致的 y 的平均变化可能是错误的。相关关系不同于因果关系。

第四部分　概　　率

13 机会是什么?

从长远观点看来,我们都是死人。

——John Maynard Keynes(英格兰。1883—1946)

1. 引言

人们一直在不造成任何损害下,不严格地谈论机会。找到工作的机会是多少?碰到某人的机会是多少?明天下雨的机会是多少?但是为了科学目的,必须给机会这个词一个确定的、清晰的解释。结果,这是困难的,数学家们为此已经奋斗了几个世纪。他们发展了几种细微而严格的机会理论;但是这些理论只覆盖了人们通常谈论机会的情况的一小部分 。本书将介绍频数论,这一理论对可以在同一条件下独立地多次重复的过程最能奏效[①]。机会游戏属于这一范畴,事实上,许多频数理论是为了解赌博问题而得到发展的。Abraham de Moivre 是这一理论的最伟大的大师之一,他是法国耶稣教徒,为了逃避宗教迫害逃到英格兰。他的著作"机会论"的部分献词刊印于下[②]。

图 1　De Moivre 给"机会论"题的献辞

尊敬的 Carpenter 爵士

我的爵士

世上有许多人被灌输了这样一种见解，即机会论具有某种助长赌博的倾向；但是，如果他们愿意浏览一下本书的大体构思，他们会很快醒悟过来；与此同时，告诉他们爵爷很高兴地支持赞助这第二版的出版，则将不无助益；如果他们的担心不是完全无根据的话，您的严肃正直和您在世界上所享有的卓越的声誉是不会允许这种赞助的。爵爷无疑明察，这一理论远非怂恿赌博，相反它把那些与机会有关的游戏的利与弊置于光天化日之下，并以此作为劝止赌博的指南…

这本机会论的另一用途是，它与数学的其它组成部分一起可作为推理艺术的合适入门：由经验知道，对掌握推理艺术来说，没有什么能比对从无可置疑的原则出发，正确演绎出来的一长串推断进行思考更有助益，对此本书提供了大量例子。

一个简单的机会游戏是抛一个硬币的赌博。抛硬币的过程可以在相同条件下独立地一次又一次地重复进行。获得头像的机会是 50%：最终大约有 50%的时刻头像向上。

取另一个例子。掷一颗骰子，得到幺点⚀的机会是 6 次中有 1 次，或 $16\frac{2}{3}\%$。解释是类似的。即当骰子在相同条件下一次又一次地重复这一基本机会过程时，最终地大约有 $16\frac{2}{3}\%$的时刻出现幺点。

当某一基本过程在相同条件下独立地一次又一次进行时，某事件的机会给出了期望该事件发生次数的百分数。

所以机会论对任何一次特定的抛硬币或掷骰子没讲得很多。但是它对长期的行为作出了明确的预测。

现在阐述关于机会的两个基本事实。

机会是在 0%与 100%之间。

如果某事件不可能发生。它发生的机会为 0%。另一个极端，如果某事件肯定要发生，那么它发生的机会为 100%。一切机会都介于这两个极端之间。

某事件的机会等于 100%减去其对立事物的机会。

例如，如果你有 45%的机会赢得游戏，那么你期望大约 45%的时刻会赢，从而你必是想到另外 55%的时刻会输。



Abraham de Moivre（英格兰，1667—1754）
Faber 蚀刻，在此复制得到大英博物馆董事会同意。

习题 A

1. 一台计算机编制了计算不同机会的程序。试将数值答案与文字描述相配（有的文字描述可用一次以上）。

数值答案	文字描述
(a)－50%	(i)这发生与不发生的可能性一样。
(b)0%	(ii)这十分可能发生，但不是一定发生。
(c)10%	(iii)这不发生
(d)50%	(iv)这能够发生，但不大可能
(e)90%	(v)这无疑会发生
(f)100%	(vi)程序中有毛病。
(g)200%	

2. 抛一枚硬币 1 000 次。可期望出现大约多少次头像？
3. 掷一颗骰子 6 000 次。可期望出现大约多少次幺点？
4. 在(取 5 张牌的)扑克游戏中，发到一副全手牌(2 张同点加 3 张 同点)的机会大约有 0.0014。如果发 10 000 副牌，大约有多少副是全手牌？

这些习题的答案在第 695—696 页上。

2. 利用终久论点

一个盒子装有编号从 1 到 10 的 10 张票：

1	2	3	4	5	6	7	8	9	10

它们全都有同样的大小、形状和质地。充分摇动盒子使票混合，并使得随机地抽取一张时，所有票子都有相等机会被抽取到。例如抽到 7 的机会是 10 次中 1 次或者 10%。设想一下，从此盒中随机地一次又一次地抽取一张票子，每次抽了一张以后再放回盒中，这样做是为了在相同条件下独立地重复这一过程。终久每一张票子大约应当象其它任何一张票子一样经常出现，10 次中大约 1 次。

例 1. 一盒装有红弹子和蓝弹子。从盒中随机地取出一颗弹子(每颗弹子都有相等机会被取出)。如果是红色的，你赢得 1 美元。如果是蓝色的，你不赢什么。你可以在下面两个盒子中选：

- A 盒装有 3 颗红弹子和 2 颗蓝弹子；
- B 盒装有 30 颗红弹子和 20 颗蓝弹子。

哪一个盒子提供较好的赢的机会，或者它们是相同的？

解　一些人看中 A 盒，因为 A 盒有较少的蓝弹子。另一些人看中 B 盒，因为 B 盒有较多的红弹子。这两种看法都是错误的。两个盒子提供相同的赢的机会，5 分之 3。

要知道为什么，设想一下，大数量次随机地从 A 盒中取弹子（每次取后放回，目的是不改变试验条件）。终究 5 颗弹子中每一颗在 5 次中将出现 1 次。因此红弹子将在 3/5 的时间出现。对 A 盒来说，你取出 1 颗红弹子的机会是 3/5 或者 60%。

现在设想大数量次从 B 盒中随机有返回地取弹子。50 颗弹子中的每颗弹子大约 50 次中将出现 1 次。但是现在有 30 颗红弹子。对 B 盒，你赢的机会是 30/50 或者 60%，这恰与 A 盒一样。

计数为比值

$$\frac{\text{红弹子数}}{\text{总弹子数}}$$

两个盒子的这一比值是相同的。De Moivre 对本例所做的解在图 2 中。

图 2　De Moivre 解

> 一个事件的概率是较大或较小，依据这一事件的可能发生机会数与这一事件的可能发生或不发生的全部机会数之比。
> 因此，如果我们构造一个分数，它的分子是一个事件藉以可能发生的机会数，而分母是藉以可能发生或不发生的全部机会数，那么这个分数将是这一事件发生的概率的恰当的表示。因此，如果一个事件有 3 次机会发生和 2 次机会不发生，分数 3/5 将恰当地表示这一事件发生的概率，并且可以取作这一概率的度量。
> 对于不发生的概率，可以同样叙说，它同样由一个分数来

度量，这一分数的分子是藉以可能不发生的机会数，而分母是它发生和不发生两者的全部机会数；因此，一个有2次机会不发生和3次机会发生的事件的不发生概率将由分数2/5来度量。

表示发生和不发生的概率的两个分数加起来，它们的和总是等于1，因为它们的分子之和等于它们的公分母：现在一个事件或者发生或者不发生是一个确定性事件，于是可以在无限大级概率概念下构想的必然性被恰当地用单位表示。〔对于“单位”一词，de Moivre 指的是数1。〕

这些事情是容易理解的，只要考虑到概率一词所含有双重意思：第一，一个事件藉以可能发生的机会数；第二它藉以可能发生或者不发生的机会数

许多问题，像例1一样，采用随机地从一盒中抽取的形式。一个典型指令是：

随机放回地从盒子

| 1 | 2 | 3 |

中抽取两张票。这要求你设想下列过程：摇动盒子，随机地（三张票都有相等机会）抽出1张票，记录这张票上所示的数，把它放回盒中，再次摇动盒子，随机地（三张票都有相等机会）第二次抽出1张票，记录这张票上所示的数，把第二张票放回盒中。对比指令是：

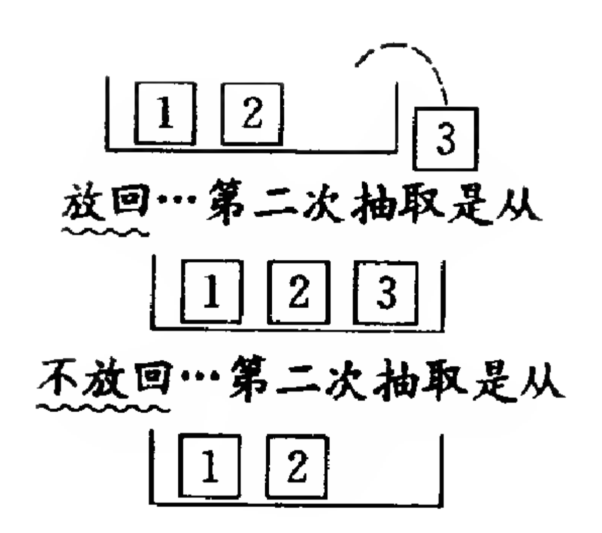
随机不放回地从盒子

| 1 | 2 | 3 |

中抽取两张票。第二个指令要求你设想如下过程：摇动盒子，随机地（三张票都有相等机会）抽取1张票，把它放在一旁，随机地（留在盒中的两张都有相等机会）抽取第二张票。见图3。

图3　放回或不放回抽取之间的差异

从盒子| 1 | 2 | 3 | 随机地作两次抽取，假设第一次抽得为 3



习题 B

1. 一盒装有 3 颗红弹子和 2 颗蓝弹子；随机放回地从盒中抽取 500 次。大约有多少红弹子可期望得到？大约有多少蓝弹子？
2. 情况与习题 1 相同，并且抽得 1 颗红弹子你赢 1 美元，抽得蓝弹子不赢什么。

   (a)如果你抽得 308 颗红弹子，你赢了多少？

   (b)如果你抽得 273 颗红弹子，你赢了多少？

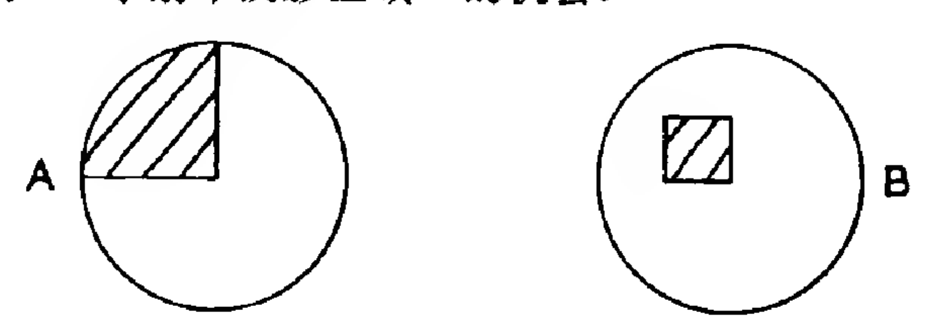
   (c)你期望能赢大约多少？
3. 如同习题 1 与 2，不过现在抽得红弹子你赢 1 美元，抽得蓝弹子你输 1 美元。你期望大约赢多少？
4. 如同习题 1、2 和 3，不过现在抽得红弹子你赢 8 美元，抽得蓝弹子你输 10 美元。你期望大约赢多少？
5. 随机放回地从下面所示的两个盒子之一抽取 100 张票。在每次抽取后，立即根据票上所示的数付给你美元。哪一个盒子较有利？为什么？



6. (a)假设一位射箭手射靶，射在任何一点的可能性与射在任何其它一点的可能性相同。问射中下面靶上阴影区域 A 的机会是多少？

   1%　10%　25%　50%　75%　90%　99%

   (b)如同(a)，求射中阴影区域 B 的机会。



这些习题的答案在第 696 页上。

3. 条件概率

本节介绍条件概率。例子涉及扑克牌。(一副纸牌有 52 张,分 4 组,每组 13 张牌。四组是:黑桃、红心、方块、梅花。每组 13 张牌:A(爱司)、K(国王)、Q(皇后)、J(宫仆)和 2 到 10。)

例 2　一副牌洗过后,最上面的两张牌面朝下地放在桌上。如果第二张牌是红心 Q,你赢 1 美元。

(a)你赢美元的机会是多少?

(b)你翻开第一张牌,这张牌是梅花 7。现在你赢的机会是多少?

解.(a)在查看任何牌之前这问题可以提问(和回答)。打赌与第二张牌有关,不是第一张;机会是 1/52。要知道为什么,考虑洗牌。它把牌置入随机排列中。红心 Q 必定插在牌中某个位置。一共有 52 个可能的位置,并且它们都是等可能的。这张 Q 插在一副牌中第二张位置,在 52 次中有 1 次机会——给你带来 1 美元。答案是 1/52。

(b)留下来有 51 张牌。它们处在随机排列中,红心 Q 是它们中一张。它放在桌上的机会为 51 次中 1 次。你的机会略有上升,为 1/51。这就是答案。

这 1/51 称为条件机会。问题给了第一张牌一个条件:它必须是梅花 7。数学家可以谈论给定第一张牌是梅花 7,或者,如果第一张牌是梅花 7,第二张牌是红心 Q 的条件概率。要强调指出一下它的对照情况(a)中的 1/52 称为无条件机会:该问题没有在第一张牌上设置条件。

习题 C

1. 随机不放回地从盒子 | 1 | 2 | 3 | 4 | 中抽取两张票子。

(a)第二次抽得的票为 4 的机会是多少?

(b)已知第一次是 2，第二次是 4 的机会是多少？

2. 在抽取有放回的情况下，重复习题 1。

3. 一颗骰子掷两次。

(a)求第二次掷出 1 点的机会。

(b)已知第一次掷出 1 点，求第二次掷出 1 点的机会。

4. 抛一枚硬币三次。

(a)求第三次抛出头像的机会。

(b)已知前二次是背面，求第三次抛出头像的机会。

5. 五张牌发自一副很好洗过的牌的最上面。

(a)求第五张牌是梅花 A 的机会。

(b)求第五张牌是梅花 K 的机会。

(c)求第五张牌是梅花的机会。

(d)求已知前四张牌是梅花，第五张牌是梅花的机会。

这些习题的答案在第 696 页上。

技术性注：(a)数学家将"第二张牌是红心 Q 的概率"记为：

P(第二张牌是红心 Q)

"P"是"概率(Probability)"的缩写。

(b)已知第一张牌是梅花 7，第二张牌是红心 Q 的条件概率记为：

P(第二张牌是红心 Q|第一张牌是梅花 7)。

垂直的一杠读做"已知"。

4. 乘法规则

本节将指出，如何通过概率相乘去计算两个事件发生的机会。

例 3. 一盒装有三张分别涂了红色(R)、白色(W)和蓝色(B)的票。

| R | W | B |

随机不放回地抽取两张票。问先抽出红色票，随后抽出白色票的机会是多少？

解　设想有许多人。每人都持有一只盒子 | R | W | B |，并且

都随机不放回地抽取两张票。在第一次抽取时大约有⅓的人得到 R ,并在盒中留下

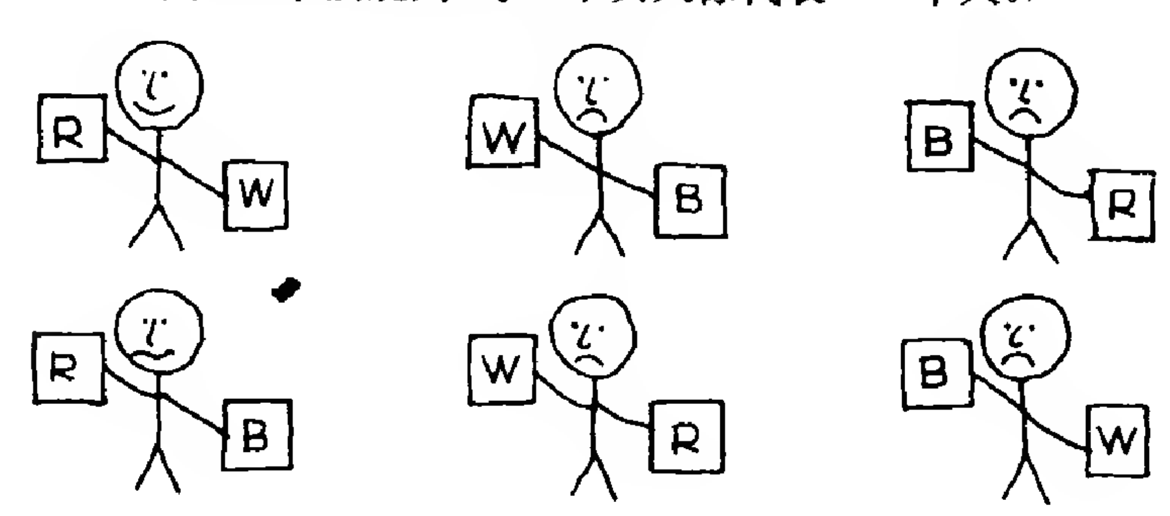
| W | B |

在第二次抽取时,这些人中大约有一半人得到 W ,因此,抽到 R W 的人的比例是

$$\frac{1}{3}\text{中的}\frac{1}{2}=\frac{1}{2}\times\frac{1}{3}=\frac{1}{6}$$

机会是 6 个中 1 个,或者 16⅔%。

图 4 乘法规则(每一木头人像代表 100 个人。)



譬如,假定开始时你有 600 人。在第一次抽取时,他们中大约有 200 人得到 R 。这 200 人中在第二次抽取时大约有 100 人得到 W ,所以有 100/600=1/6 的人第一次抽得红色票,而且随后又抽得白色票。参见图 4;抽到 R W 的人在左上方。

统计学家通常以相反的顺序进行机会乘法:

$$\frac{1}{3}\times\frac{1}{2}$$

他们的理由是:1/3 属于第一次抽取,而 1/2 属于第二次。

例 3 中的方法称为乘法规则

两件事一起发生的机会等于第一件发生的机会乘以已知第一件发生的情况下第二件发生的机会

例 4. 两张牌发自一副洗匀的牌的最上面。第一张牌是梅花 7 和第二张牌是红心 Q 的机会是多少?

解　这与例 3 中拥有一个大得多的盒子的情况相同。第一张牌是梅花 7 的机会是 1/52。已知第一张牌是梅花 7,第二张牌是红心 Q 的机会为 1/51。两张牌同时得到的机会是

$$\frac{1}{52}\times\frac{1}{51}=\frac{1}{2\ 652}$$

这是一个很小的机会:大约 10 000 次中有 4 次,或者 1%的 0.04。

例 5. 从一副洗好的牌中发两张。两张都是 A 的机会是多少?

解　第一张牌是 A 的机会等于 4/52。已知第一张牌是 A,余下的 51 张牌中有 3 张 A:第二张是 A 的机会等于 3/51。两张牌都是 A 的机会等于

$$\frac{4}{52}\times\frac{3}{51}=\frac{12}{2\ 652}$$

这大约是 200 次中 1 次,或者 1%的 1/2。

例 6. 一枚硬币抛两次。一次头像随后一次背面的机会是多少?

解　第一次抛出头像的机会等于 1/2。不管第一次抛出什么,第二次抛出背面的机会都等于 1/2。所以背面接着头像的机会等于

$$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

习题 D

1. 一颗骰子掷两次,得到两个幺点的机会是多少?
2. 将一副牌洗过后发出三张。第一张是 K、第二张是 Q 和第三张是 J 的机会是多少?
3. 一枚硬币抛三次。求两次头像随后一次背面的机会。
4. 一颗骰子掷六次。你在两种选择中可选一种:

(i)如果至少出现一个幺点,则赢 1 美元;

(ii)如果所有次中显示一个幺点，则赢 1 美元。

哪一种选择提供较好的赢的机会？或者它们有相同的机会？说明之。

5. “一条九____鞭可以用于惩罚国家____，* ，不过很少这样做。”将一枚硬币抛两次，把两次结果依次填入空白。填对的机会是多少？

6. 1986 年美国人口中大约有 51%是女性。另外，人口中大约 12%是 65 岁和 65 岁以上[3]。正确还是错误，并解释之：年纪在 65 岁和 65 岁以上的妇女构成人口的百分数为

$$12\%\text{中的 }51\% = 0.51\times 12\% \approx 6\%$$

这些习题的答案在第 696 页上。

5. 独立性

本节引入独立性概念，这一概念将多次在本书的余下部分中使用。

> 两件事是独立的，如果给定第一件事，无论它的结果是什么，第二件事物的机会都一样。否则，它们是不独立的。

例 7. 某人抛一枚硬币两次。如果第二次抛得头像，你赢 1 美元。

(a)假如第一次抛得头像，你赢 1 美元的机会是多少？

(b)假如第一次抛得背面，你赢 1 美元的机会是多少？

(c)这两次抛硬币独立吗？

解：如果第一次抛得头像，第二次有 50%机会得到头像。如果第一次抛得背面，它的机会仍然等于 50%。这就是说，不管第一次抛得什么，第二次抛得各种结果的机会保持不变。那就是独立性。

例 8. 一盒装有三张编号为 1、2、3 的黑色票和三张编号为 1、2、3 的白色票。从中随机地抽取一张，要求你猜这张票上的号码

---

* 硬币的头像(Head)又可译为“元首”，背面(Tail)又可译为“尾”。因此相当于随机放回地将“元首”“尾”两词填入习题 5 的两个空白。——译者注。

[■] [■] [■] [1] [2] [3]

(a)当票从盒子中取出时,你瞥了一眼。你没有能够辨认出票上的号码,但是看到此票是黑色的。票上的号码为 2 的机会是多少?

(b)除了票是白色外,其它情况同(a)。

(c)颜色与号码独立吗?

解:如果票子是黑色的,3 次中有 1 次机会出现号码 2。如果票子是白色的,完全类同。由此知道颜色不改变号码的机会。因而颜色与号码是独立的。

例 9. 从盒子

[■] [■] [■] [1] [1] [2]

中随机地抽取一次。颜色与号码独立吗?

解:如果票子是黑色的,3 次中有 2 次机会出现号码 2。如果此票是白色的,出现号码 2 的机会减少到 3 次中 1 次。在已知颜色下,号码出现的机会变了。因而号码与颜色是不独立的。

例 10. 随机放回地从盒子

[1] [1] [2] [2] [3]

中抽取两次。

(a)假定第一次抽得[1]。第二次抽得[2]的机会是多少?

(b)假定第一次抽得[2]。第二次抽得[2]的机会是多少?

(c)两次抽取独立吗?

解:不管第一次抽得的是[1]还是[2],还是任何其它什么,在第二次抽取时得到[2]的机会保持相同——5 次中 2 次,或者 40%。理由是:第一张票子放回盒中,所以第二次抽取总是从同一盒子[1] [1] [2] [2] [3]中作出的。抽取是独立的。

例 11. 如在例 10 中一样,但是两次抽取是在不放回条件下进行的。

解：如果第一次抽得[1]，那么第二次是从盒子|[1] [2] [3] [4]|中抽取的。第二次抽得[2]的机会是50%。另一方面，如果第一次抽得[2]，那么第二次是从盒子|[1] [1] [2] [3]|中抽取的。现在第二次抽得[2]的机会只有25%了。抽取是不独立的。

> 随机有放回抽取时，各次抽取是独立的。不放回抽取时，各次抽取是不独立的。

各次抽取的独立性指的是什么意思？为了回答这个问题，集中关注由一次抽取以确定输赢的打赌：比如说，抽取次数是3次或更多。那么，无论其它抽取的结果是什么，赢得这场打赌的条件概率保持不变。

> 如果两个事件是独立的，那么这两个事件都发生的机会等于它们各自无条件概率的乘积

例12. 一盒装有分别涂上红、白和蓝色的三张票。

|[R] [W] [B]|

从中随机有放回地抽取两张。先抽得红色票，随后抽得白色票的机会是多少？

解：抽取是独立的，因此所求机会是

$$\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$$

把本例与例3作比较；答案不同。这与独立性有关。这次较容易，因为你不需要计算条件概率。对于独立事件，乘法规则需要做的事情较少。

习题 E

1. 对下列盒子中每一个，说出颜色与号码是独立的还是不独立的。

(a) 1 2 2 **1** **2** **2**

(b) 1 2 **1** **2** **1** **2**

(c) 1 2 3 **1** **2** **2**

2. 一枚硬币抛三次。

   (a)得到三次头像的机会是多少?

   (b)没有得到三次头像的机会是多少?

   (c)至少得到一次背面的机会是多少?

   (d)至少得到一次头像的机会是多少?

3. 每周你买一张彩票,它提供百万分之一的赢的机会。尽管你坚持买了十年之久,你从未赢过的机会是多少?

4. 随机不放回地从盒子 1 2 3 4 中抽取两次。第一张票遗失了,而且没有人知道它上面写的号码。在这情况下两次抽取是独立的,这一断言是否正确,说明之。

5. 假设某一班级有

   • 80%男生和20%女生;

   • 15%一年级生和85%二年级生

   (a)班级中二年级女生的百分数能与________一样小。

   (b)这一百分数能与________一样大。

6. (难)在某次心理学实验中,给每一位实验对象三张面朝下的牌。实验对象从这三张牌中取一张;并且随机地从一副单独的完整的扑克牌中取一张。如果这两张牌同花,那么实验对象赢得一件奖品。试问赢的机会是多少?或这能由已知的信息决定吗?

这些习题的答案在第696—697页上。

6. Collins 案件

People v. Collins 是一桩诉讼案,在这案件中有一个重要的统计问题。一黑人男子与一白人女子被指控抢劫。法庭叙述的事实如下[4]:

1964年6月18日上午11:30左右,Juanita Brooks 太太购物以后,沿着洛杉矶城圣·彼得罗区一条小巷走回家。在身后她拖着

一个装有食品杂货的柳条编的小筐，并且在这包东西的上面放有她的钱包。她拄着一根手杖。当她弯下腰去拾起一只空纸盒时，她突然被一个她既未看到又未听到走近的人推倒在地。她由于跌倒而大吃一惊，并且感到有点疼痛。她设法仰起头并看到一个从出事地点跑开的年青妇女。据 Brooks 太太说，后者的体重好像有 145 磅左右，穿着“有点暗”，且头发是“深亚麻色与淡亚麻色之间”，但是这比被告 Janet Collins 在庭审时的头发淡一些。在事件刚过之后，Brooks 太太马上就发现她装有 35 美元到 40 美元之间的钱包丢失了。

大约在抢劫的同一时刻，居住在这条小巷一端街上的 John Bass 正在他屋前给草坪浇水。从小巷里传来的“大量喊叫声”引起了他的注意。当他向那一方向看时，他看见一个妇女跑出这条小巷，并进入一辆停在马路对面的黄色汽车。他不能说出汽车的式样。汽车立刻开动，并且远远地绕过另一辆停着的机动车，从而在离 Bass 不到 6 英尺的狭窄的街道上驶过。此时 Bass 看到，这辆车是由一个有小胡子和络腮胡子的黑人男子驾驶的。在庭审时 Bass 认定被告为黄色轿车的驾驶员。尽管如此，仍试图对他的辨认提出异议，因为他承认，袭击 Brooks 后不久在警察局对一列嫌疑犯辨认时，当时被告没胡子，他声称不能肯定辨认。

Bass 在他的证词中，描述从小巷中跑出的妇女是一个白种人，身高略超过五英尺，正常身材，她的头发是深亚麻色马尾巴型，穿深色外套。他进一步证实，她的马尾巴发型与珍妮特 1964 年 6 月 22 日在警察局拍的照片“确实相像”。

随后检察官让当地州立学院的一位数学讲师讲解乘法规则，他没有对独立性，或者对条件概率与无条件概率之间的区别赋于很多关注。在这篇证言之后，检察当局假设了如下的机会：

黄色轿车	1/10	有亚麻色头发的妇女	1/3
有小胡子的男人	1/4	络腮胡子的黑人男子	1/10
梳马尾巴发型的妇女	1/10	车中一对不同种族	1/1000

把这些机会全部相乘，得出 12 000 000 分之一。按照检察当局的说法，这一方法给出了“任何[另外]一对具有被告那与众不同的特征”的机会。如果没有另外一对具有这些特征，被告是有罪的。

陪审团宣判有罪。在上诉时，加利福尼亚最高法院推翻了陪审团的裁决。它发现没有依据支持这六个机会的假设值。此外，这些值都是当作无条件概率提出的。那位数学讲师本来应该解释，它们相乘的依据是独立性。再者并不存在支持独立性假设的依据。相反，某些因子显然是不独立的——如“络腮胡子的黑人男子”与“车中的一对不同种族”。

有一个更加具有基本性的反对理由。像乘法规则这样的概率运算是为了处理机会游戏发展起来的，在这种游戏中基本过程可以在同一条件下独立地重复进行。事实上，按照频数论，一个事件的概率是这事件在长期进程中发生次数的百分率。检察官试图把这一理论用于唯一一个事件：1964 年 6 月 18 日上午 11：30 或者发生过——或者没有发生过——的某件事。在这新的含义下，机会指的是什么？检察官有责任回答这一问题并证明这理论适用于他的情况[5]。

7. 复习题

1. 下列结论正确还是错误，并解释之：
   (a) 如果某事物具有概率 1000%，它肯定会发生。
   (b) 如果某事物具有概率 90%，那么每当它的对立事物发生一次时，它可望发生大约 9 次。
2. 在(取 5 张牌的)扑克游戏中，发到一付对子或更好的牌的机会有 49.8%。一手牌，如果一付对子也没有，称为破产。问得到破产的机会是多少？
3. 从一副洗匀的牌的最上面发两张牌。有如下两种选择：
   (i) 如果第一张牌是 K，赢 1 美元
   (ii) 如果第一张牌是 K 和第二张牌是 Q，赢 1 美元。哪一种选择

较好？还是它们是等价的？简短地解释之。

4. 从一副洗好的牌中。要求你每次发一张牌直到A出现。你已经发了10张牌，但是仍然没有见到A。问：发第11张牌时你得到A的机会是多少？
5. 发了一手牌(五张)。求前4张是A和第5张是K的机会。
6. 随机地从下面盒子中抽取一张票。问：颜色与号码是否独立？

1	1	8	1	1	8

7. 一副洗好的牌，将最上面两张牌面朝下地放在桌上。下列结论是否正确？说明之。

   (a)第一张牌为A的机会是13次中1次。

   (b)第二张牌为A的机会是13次中1次。

   (c)两张牌都是A的机会是1/13×1/13。
8. 一枚硬币抛六次。两种可能的结果序列是

   (i) H T T H T H　　(ii)H H H H H H(硬币必须按给出的次序出现H或T：H＝头像，T＝背面。)下面哪一个结论是正确的？解释之[⑥]。

   (a)序列(i)更可能。

   (b)序列(ii)更可能。

   (c)两个序列同样可能。
9. 一颗骰子掷三次。三次都掷出3点或3点以上的机会是多少？
10. 一颗骰子掷四次。求得到四个6点的机会。求没有得到四个6点的机会。
11. 从下面两个盒子的一个中随机放回地抽取100次。每一次抽取，付给你票上所示数的美元。哪一个盒子较有利？用这盒子你期望赢得多少？

    (i) | 1 | 1 | 5 |　　(ii) | 1 | 5 |

8. 小结

1. 机会的频数论大多直接用于能够独立地在同样条件下一次又一次重复的机会过程。

2. 某事的机会是指，在基本过程一次又一次重复时，这事可望发生的次数的百分率。

3. 机会在0%与100%之间。用0%表示不可能性。100%表示必然性。

4. 某事件的机会等于100%减去它的对立事件的机会。

5. 两件事同时发生的机会等于第一件事发生的机会乘以已知第一件事发生情况下第二件事发生的条件机会。

6. 如果不管第一件事发生结果如何，第二件事发生的机会都保持相同，那么这两件事是独立的。

7. 如果两件事是独立的，这两件事都发生的机会等于它们的无条件概率的乘积。

8. 机会的数学理论适用于某些场合。在另外一些场合使用可能会导致荒谬的结果。

14
再谈机会

对某些表面很简单的机会问题，思想上容易陷于相信，它们的解只要凭藉天赋机智的力量就可获得；往往可以证明事情并不如此，而且由此引起的错误也并不罕见，这足以证明，一本教会人们把真理和看起来非常接近真理的事物区分开来的这类书，可视为适当推理的帮手。

——Abraham De Moivre（英格兰 1667—1754）[①]

1. 列举状态法

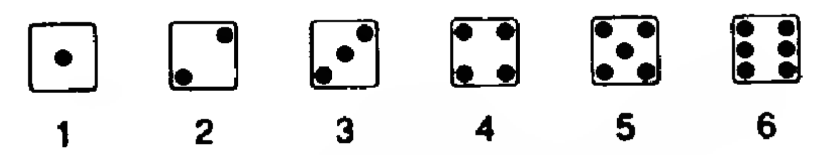
一位概率论专家是一位专门研究计算复杂事件概率的数学家。廿世纪两位最主要的概率论专家是 A. N. Kolmogorov（俄罗斯 1903—1987）和 P. Levy（法国 1886—1971）。他们发展的方法超出本书的范围，但是我们可以考虑由较早期的数学家所发展起来的更为基本的方法。

在试图演算机会时，有时列举出一个机会过程可能出现的全部状态是十分有用的。如果这样做太困难，那么写出其一小部分典

型的状态是一个良好的开端。求出状态的总数往往是可能的，而且它能像列举出它们全体一样有用。

例 1. 掷一颗骰子。出现一个偶数点(2、4 或 6)的机会是多少？

解　这里机会过程由抛骰子构成。骰子落下时有六种可能状态：



这六个状态中每一个都有相同的机会，6 分之 1。有 3 个状态得出偶数，因此机会是 6 分之 3。

例 2　掷一对骰子。得出点数之和为 2 的机会是多少？

解　这里机会过程由抛两颗骰子构成。每颗骰子显示的点数是什么状态呢？为了区分它们，设想一颗骰子是白色的，而另一颗是黑色的。骰子落下时的一种状态是



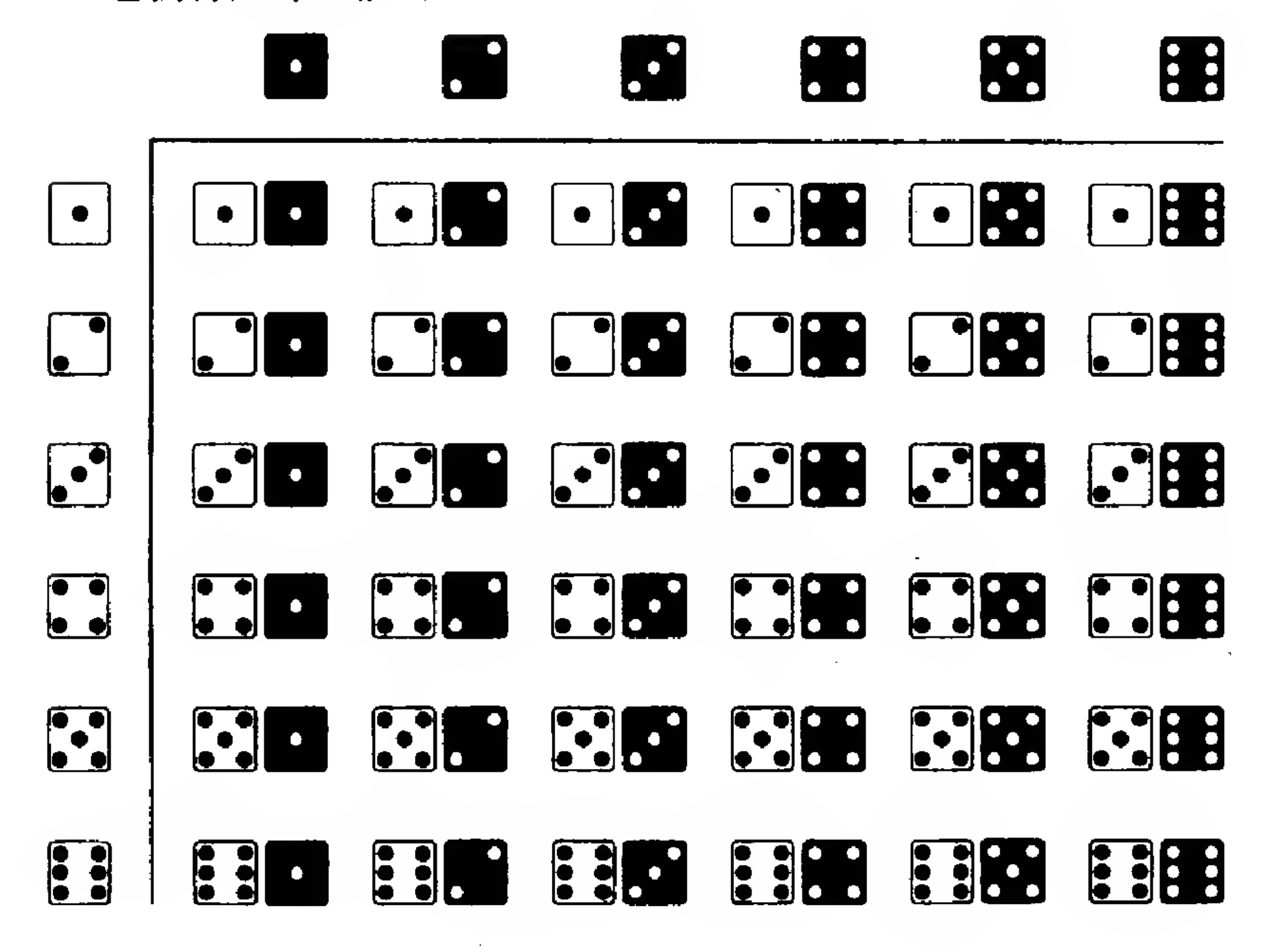
这意味着白色骰子显示 2 点，黑色骰子显示 3 点：点数之和为 5。

两颗骰子落下时总共有多少状态呢？首先，白色骰子能够以 6 种状态中任何 1 种落下：

譬如当白色骰子显示　时，黑色骰子仍有 6 种状态可以落下呈现：

现在两颗骰子能够出现的可能状态有 6 个。这些状态表示在图 1 的第一行中。类似地，第二行表示白色骰子显示　时两颗骰子可能呈现的另外 6 个状态，等等。图中表示出两颗骰子出现的可能状态有 6×6＝36 个。它们都是等可能的，所以每一个状态都有 36 次中 1 次机会。只有一个状态得出和为 2 点：　。因此它的机会是 1/36。那就是答案。

图 1 掷两颗骰子。骰子落下时有 36 种状态，展示在图的主体中，它们都是等可能的



通常有几种方法获得与机会有关的问题的解。例如在图 1 中，36 个结果中每一个的机会也可以用乘法规则得出：$1/6 \times 1/6 = 1/36$。

例 3. 掷一对骰子。得出点数和为 4 的机会是多少？

解：看图 1。有 3 个状态得出和为 4 点：



所以机会是 36 分之 3。这就是答案。

三颗骰子的情况怎样？像图 1 一样的三维图是多了一点累赘。但是类似的推理仍然可以使用。例如，十七世纪的意大利赌徒经常在三颗骰子掷出的点数和上打赌。他们深信，掷出的点数和为 9 的机会应该等于掷出的点数和为 10 的机会。例如他们说，掷出点数和为 9 的一个组合是：在一颗骰子上出现 1 点，另一颗上出现 2 点，第三颗上出现 6 点。这可以缩写成“1 2 6”。对于和为 9 点一共

有六个组合：

1 2 6　1 3 5　1 4 4　2 3 4　2 2 5　3 3 3

类似地，对于和为 10 点他们找到六个组合为：

1 4 5　1 3 6　2 2 6　2 3 5　2 4 4　3 3 4

于是，赌徒们争辨，按理 9 和 10 应当有相同的机会。但是经验表明，10 发生的次数比 9 多一点。

他们请 Galileo（伽里略）帮助解决这一矛盾，他推理如下。将一颗骰子染成白色，一颗灰色，另一颗黑色，这样它们可以区分。这不会影响机会。现在三颗骰子落下会出现多少种状态？白骰子能处于 6 种状态。对应于这 6 种状态的每一种灰骰子能处于 6 种状态，使得这两颗骰子有 6×6 种可能状态。并且对应这些可能状态中每一种。黑骰子仍然有 6 种状态。总共三颗骰子可处的状态有 $6\times6\times6=6^3$ 种。（四颗骰子将有 6^4 种；五颗骰子将有 6^5 种，等等。）

现在，三颗骰子落下的状态很多，有 $6^3=216$ 个。但是 Galileo 还是坐下来列举出它们，随后，他全面检查了他的表并算出点数和为 9 的状态数；求得为 25。同时他也求得点数和为 10 的状态数 27。他推断出，掷出 9 的机会是 $25/216=11.6\%$，而掷出 10 的机会是 $27/216=12.5\%$。

赌徒们的讨论没有认真对待骰子落地时的不同状态。例如，关于 9 的一组 3 3 3 只对应骰子落地时一种状态：



但是，对 10 来说，3 3 4 这组对应着骰子落地时的三种状态：



赌徒们的讨论可以如表 1 那样予以校正。

表 1 经 Galileo 校正后的赌徒讨论给出的三颗骰子得到 9 或 10 点的机会。

9 点的三数组	每组掷出的状态数	10 点的三组数	每组掷出的状态数
1 2 6	6	1 4 5	6
1 3 5	6	1 3 6	6
1 4 4	3	2 2 6	3
2 3 4	6	2 3 5	6
2 2 5	3	2 4 4	3
3 3 3	1	3 3 4	3
和	2 5	和	2 7

习题 A

1. 参见图 1，作出一张骰子掷出点数之和为 5 的状态表。两颗骰子掷出点数和为 5 的机会是多少？
2. 一对骰子掷 1 000 次。那一种点数和出现次数最多？那一种点数和出现次数最少？
3. 随机放回地从盒子 [1] [2] [3] [4] [5] 中抽取两次。绘制一张如图 1 那样的图来表示全部可能结果。有多少个结果？两次抽取所得之和等于 6 的机会是多少？

这些习题的答案在第 697 页上。

2. 加法规则

本节涉及两件指定事中至少有一件发生的机会：或者第一件发生、或者第二件发生、或者两件都发生。两件事同时发生的可能性原来是一件较复杂的事情，但是这种复杂性有时可以排除。

> 两件事是互不相容的，如果一件事发生制止另一件事的发生。

例 4. 从一副很好洗过的牌的最上面发一张牌。这张牌可能是



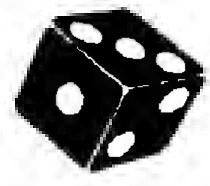
Galileo(伽里略)(意大利　1564—1642)

来自:Wolff－Leavenworth 收藏在 Syracuse 大学的 George Arents 研究图书馆

红心,或者可能是黑桃,这两种可能互不相容吗?

解　假如这张牌是红心,它就不可能是黑桃。这两种可能性是互不相容的。

例 5. 某人掷一对骰子。一颗是带黑点子的白色骰子,另一颗是带白点子的黑色骰子。

下面的结果是互不相容的吗?

- 白色骰子呈现幺点
- 黑色骰子呈现幺点

解　骰子能够呈现 。两个结果不是互不相容的,因为白色幺点不制止黑色幺点。(如图 1 所示,两个结果都发生的机会是 36 次中 1 次。)

现在我们可以阐述计算机会的一般准则。它称为**加法规则**。

> 为求两件事中至少有一件发生的机会，检查一下它们是否互不相容，如果是的，把它们的机会相加。

例 6. 从一副洗匀的牌的最上面发一张牌。这张牌是红心的机会为 4 分之 1。是黑桃的机会为 4 分之 1。它是一张实力较强的花色（红心或黑桃）牌的机会是多少？

解　问题是要求下列两件事中一件发生的机会：

- 这张牌是红心；
- 这张牌是黑桃。

如在例 4 中一样，如果这张牌是红心，那么它就不可能是黑桃：这两件是互不相容的事件。因此，机会相加是合法的。得到一张强花色牌的机会是 $1/4+1/4=1/2$。（验证：有 13 张红心和 13 张黑桃，所以一副牌中有 $26/52=1/2$ 是强花色的。）

盲目地把机会相加可能会得出错误的答案，这是由于重复计入两件事同时发生的机会。对于互不相容事件不存在重复计入问题。

例 7. 某人掷一对骰子。正确还是错误：至少得到一个幺点的机会是 $1/6+1/6=1/3$。

解　这是错误的。如在例 5 中一样，设想一颗骰子是白色的，另一颗是黑色的。白色骰子出现幺点不能制止黑色骰子出现幺点。机会相加会重复计算两颗骰子都出现幺点的机会。

看图 1：白色骰子显示 [·] 有 6 种状态；黑色骰子显示 [■] 有 6 种状态；但是至少得到一个幺点的状态数不是 6+6。相加就重复计入了图的左上角顶端的结果 [·][■] 。至少得到一个幺点的机会是 $(6+6-1)/36=11/36$。

自然会问，“我在何时相加机会，又在何时相乘？”答案依赖于你正试图求的机会。在要进行任何运算之前，有两步要做：

· 试图把问题有关的机会过程具体化为掷骰子、发牌或者诸如此类的事。

· 确定需要求出机会的事件。

随后，试图将这一事件与一些已经知道其机会的较简单事件联系起来。你可能想计算这些较简单事件中至少有一件发生的机会。或者，你可能想计算这些较简单事件全都发生的机会。

· 当你想求至少一件事发生的机会时，假如这些事互不相容，你就把它们的机会加起来。

· 当你想求所有这些事物全都发生的机会时，就做乘法。

有时你也许不能去加或者去乘，于是需要更多的思路（如下一节中所阐述的）。存在许多与机会有关的难题，但是请不要耽心。本书这一部分的主要目的是给你指出两个要回避的陷阱——盲目地相加机会，或者盲目地相乘机会。

技术性注. 两个幺点的机会是 1/36，所以例 7 中的机会可以计算为

$$\frac{1}{6}+\frac{1}{6}-\frac{1}{36}=\frac{11}{36}$$

然而，如果骰子掷了 3 次，至少得到一个幺点的机会不是

$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}-(\frac{1}{6})^3。$$

（考虑掷 12 次！）这类问题在下一节将以另一方法解。

习题 B

1. 50 个儿童出席一次聚会，会上供应小甜饼和冰淇淋。12 个儿童吃小甜饼，17 个儿童吃冰淇淋。正确还是错误：29 个儿童吃小甜饼或者冰淇淋，简短地说明之。
2. 从一副洗匀的牌的最上面发两张牌。你有一次如下的选择：
   (i) 如果第一张牌是 A 或者第二张牌是 A，你赢得 1 美元；
   (ii) 如果至少两张牌中有一张是 A，你赢得 1 美元。

   哪一种选择较有利？还是它们是同样有利？简短说明之。
3. 掷两颗骰子。第一颗出现[·]的机会是 1/6。第二颗出现[·]的机会是 1/6。正确还是错误：第一颗出现[·]或者第二颗出现[·]的机会等于 1/6＋1/6。

简短地说明之。

4. 一盒装有编号从 1 到 10 的 10 张票。随机放回地从该盒中抽取 5 次。正确还是错误：至少 1 次得到 [7] 的机会是 10 分之 5。简短地说明之。

5. 随机地从一盒中抽取一个数。这数等于或小于 10 的机会有 20%。它等于或大于 50 的机会有 10%。正确还是错误：得到一个数在 10 与 50 之间（不包括 50）的机会是 70%。简短地说明之。

6. 事件 A 的无条件概率是 1/2。事件 B 的无条件概率是 1/3。说出下列每一个结论是正确还是错误的，并简短地说明之。

(a)A 与 B 都发生的机会是 1/6。

(b)如果 A 与 B 独立，它们都发生的机会是 1/6。

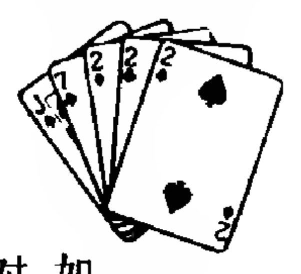
(c)如果 A 与 B 互不相容，它们都发生的机会是 1/6。

(d)A 或 B 中至少发生一件的机会是 5/6。

(e)如果 A 与 B 独立，它们中至少发生一件的机会是 5/6。

(f)如果 A 与 B 互不相容，它们中至少发生一件的机会是 5/6。

7. 在（抽取 5 张的）扑克牌游戏中，约有 2.1% 的机会发到 3 张同点牌，如



约有 4.6% 的机会发到两对，如



问得到或者三张同点或者两对的机会是多少？

8. 对于一副洗过的牌。分别说出下列结论正确还是错误，并简短地解释之：

(a)最上面一张牌是梅花 J 的机会等于 1/52。

(b)最下面一张牌是方块 J 的无条件机会等于 1/52。

(c)最上面一张牌是梅花 J 或者最下面一张牌是方块 J 的机会等于 2/52。

(d)最上面一张牌是梅花 J 或者最下面一张牌是梅花 J 的机会等于 2/52。

9. 一大群人在为能到费城去度一个全部免费的周末而进行竞赛。主持人给每一个竞争者一副完全洗过的扑克牌。竞争者从这副牌的最上面取两张牌，如果第一张牌是红心 A 或者第二张牌是红心 K，那么他就赢得去费城度

周末。

(a)第一张牌是红心 A 的所有竞争者都被要求向前走一步。多少比例的竞争者做此事?

(b)竞争者回到原先的位置。然后,要求得到第二张牌为红心 K 的竞争者向前走一步。多少比例的竞争者做此事?

(c)竞争者中有人两次向前走一步吗?

(d)正确还是错误,并解释:赢得到费城度周末的机会是 1/52+1/52。

10. 一大群人在为能到费城去度一个全部免费的周末而进行竞赛。主持人给每一个竞争者一副完全洗过的扑克牌。竞争者从这副牌的最上面取两张牌,如果第一张牌是红心 A 或者第二张牌是红心 A,那么他就赢得去费城度周末。

(a)第一张牌是红心 A 的所有竞争者都被要求向前走一步。多少比例的竞争者做此事?

(b)竞争者回到原先的位置。然后,要求得到第二张牌为红心 A 的那些竞争者向前走一步。多少比例的竞争者做此事?

(c)竞争者中有人两次向前走一步吗?

(d)正确还是错误,并解释:赢得到费城去度周末的机会是 1/52+1/52。

这些习题的答案在第 697—699 页上。

3. Chevalier De Méré 悖论

十七世纪,法国的赌徒们常常在这样的事件上打赌,即掷 4 次骰子,至少有一次出现幺点;幺点是[•]。在另一赌博中,他们在这样的事件上打赌,即一对骰子掷 24 次,至少有一次出现双幺点;双幺点是一对骰子呈现[•] [•]。

Chevalier de Méré,当时法国的一名贵族,认为这两个事件是等可能的。对于第一种赌博,他这样推断:

- 一颗骰子掷一次,我有 1/6 机会得到幺点。
- 所以掷 4 次时,我有 4×1/6=2/3 的机会至少得到 1 个幺点。

对于第二种赌博,他的推断是类似的:

- 一对骰子掷一次,我有 1/36 的机会得到双幺点。

· 所以掷 24 次时，我一定有 24×1/36 的机会得到至少一个双幺点。

根据这种论点，两个机会相同，即 2/3。但是经验表明，第一个事件比第二个事件的可能性大一些。这个矛盾已经成为众所周知的 Chevalier de Méré 悖论。

De Méré 向哲学家 Blaise Pascal 请教这个问题，Pascal 在他的朋友 Pierre de Fermat 的帮助下解决了这个问题，Fermat 是一位法官和议员，由于他在工作之余所做的数学研究，至今为人们所记得。Fermat 发现，de Méré 对不是互不相容的事件用了加法规则。毕竟，第一次和第二次掷骰子时有可能两次都得到幺点。事实上，把 de Méré 的论证稍加推进，将得出一颗骰子掷 6 次得到一个幺点的机会是 6/6，或者 100%。必定在什么地方出了错误。

现在，问题是如何正确计算这些机会。Pascal 与 Fermat 用一种典型的间接数学推理解决这个问题——这类推理总是使非数学工作者感到有点受骗。当然，像 Galileo（第一节）那样直接着手，可能很容易陷入困境：一颗骰子掷 4 次，有 $6^4=1296$ 个结果困扰着；一对骰子掷 24 次，有 $36^{24}\approx 2.2\times 10^{37}$ 个结果。

不幸的是，Pascal 与 Fermat 之间的谈话已失去了历史记载。但是在这里给出一份复制品②。

Pascal　首先让我们看第一种赌博。

Fermat　好。赢的机会很难计算，所以让我们先计算对立事件：输的机会。于是

$$赢的机会=100\%-输的机会$$

Pascal　同意。当掷了 4 次没有出现一个幺点时，赌徒输了。不过你将如何计算这些机会呢？

Fermat　看来很复杂。让我们从掷一次开始。第一次没有掷出幺点的机会是多少呢？

Pascal　必须出现 2 点到 6 点中某一个，所以机会是 5/6。

Fermat　这是事实。现在头两次都没有掷出幺点的机会是多少呢？



Blaise Pascal（法国，1623－1662）
来自 Wolff-Leavenworth 收藏在 Syracuse 大学的 George Arents 研究图书馆



Pierre de Fermat（法国，1601－1665）
来自加州大学伯克利分校的 Oeuvres Complètes 图书馆

Pascal　我们可以用乘法规则。第一次没有掷出幺点和第二次没有掷出幺点的机会等于 $5/6\times5/6=(5/6)^2$。毕竟，每次掷骰子都是独立的，难道不是这样的吗？

Fermat　掷三次又怎么样呢？

Pascal　这看上去象是 $5/6\times5/6\times5/6=(5/6)^3$。

Fermat　对。现在掷四次又怎么样呢？

Pascal　必定是 $(5/6)^4$。

Fermat　是的，那大约为 0.482，或者 48.2%

Pascal　因此输的机会有 48.2%。现在

$$\text{赢的机会}=100\%-\text{输的机会}$$
$$=100\%-48.2\%=51.8\%$$

Fermat　这就解决了第一种赌博。赢的机会稍微超过 50%。现在第二种赌博又怎样呢？

Pascal　好的，在一对骰子掷一次时，36 次中有 1 次机会得到双幺点，而 36 次中有 35 次机会没有得到双幺点。由乘法规

则，在一对骰子掷 24 次中，没有得到双幺的机会必定是

$$(35/36)^{24}$$

Fermat　大约是 50.9%，我们求得了输的机会。现在

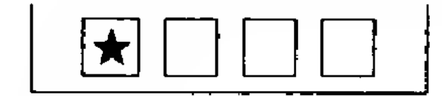
$$赢的机会=100\%-输的机会$$
$$=100\%-50.9\%=49.1\%$$

Pascal　是的，这略小于 50%。全在这里了。这就是为什么在第二种赌博中你所赢的常常会比第一种赌博少一点的原因。但是你必须大量地掷骰子才能看出这种差异。

这一例子阐述了计算机会的一种很好的策略：假如一事件的机会很难求，可以尝试一下求出其对立事件的机会；然后从 100% 中减去它。这很有用，因为对立事件的机会也许较容易计算。

习题 C

1. 一盒装有 4 张票，一张标上一颗星号，其余三张均是空白的：



随机有放回地从盒中抽取两次。

(a)第一次抽取得到一张空白票的机会是多少？

(b)第二次抽取得到一张空白票的机会是多少？

(c)第一次抽取得到一张空白票和第二次抽取也得到一张空白票的机会是多少？

(d)两次抽取都没有得到星号票的机会是多少？

(e)两次抽取中至少有一次得到星号票的机会是多少？

2. (a)一颗骰子掷 3 次。至少得到一个幺点的机会是多少？

(b)同上。但掷 6 次。

(c)同上。但掷 12 次。

3. 一对骰子掷 36 次。至少得到一次双幺的机会是多少？

4. 据 de Moivre 说，在十八世纪的英格兰，人们玩一种类似于近代轮盘赌的游戏。叫做“Royal Oak（皇家橡树）。”在桌上有 32 个“点”或编了号码的球囊，以这样一种方式抛球，使它有相等的机会，32 分之一，落入每一个球囊。

如果你在某个点处下一英磅赌注而球来到此点，你收回你的赌注并同时

赢得 27 英磅。如果你下注的点不来球，你就输掉你的一磅。游戏者（或者像 de Moivre 称呼的“冒险家”）抱怨游戏是不公平的，如果他们下注的点来球，他们应当赢得 31 磅。（他们是对的；参阅 315 页上的习题 7。）de Moivre 继续说道：

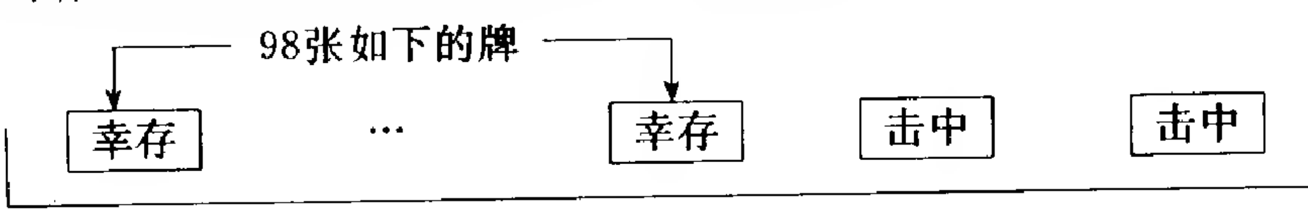
球的庄家坚持认为他们没有理由可以抱怨；因为他保证，任何一个特定的球点都应该在抛 22 次球中出现：他表示愿意就这事向人打赌，并且当有人提出要求时，他真的与人打了赌。31 比 1 的赔率与任一[点]在抛 22 次球中都会出现球，这两者之间表面上的矛盾使得冒险家们感到困惑，他们开始认为优势是在他们一边：为此理由，他们赌下去且不断地输。

抛 22 次球中，譬如说，点 17 出现的机会是多少呢？（球的庄家按同额下注打赌，因此，如果机会超过 50%，他也就证明了这样可得益。）



5. Len Deighton 在他的小说轰炸机中辩称，第二次世界大战中的飞行员在每次执行轰炸任务时有 2% 的机会被击中，因此在 50 次轰炸任务中他“数学上肯定”会被击中：$50\times2\%=100\%$。这是一个充分的论据吗？

提示：为计算机会，必须知道所考虑的情况与一个机会游戏多么相象。这里的比拟是随机有放回地从下面盒子中抽取 50 次。50 次战斗任务幸免于难好像从盒中抽到 50 次“幸存”牌。机会是多少？



这些习题的答案在第 699 页上。

4. 现实的骰子是公平的吗？

根据 Galileo 的说法(第 1 节)，在掷一颗骰子时，它的六个面的任何一面出现都是等可能的。Galileo 所考虑的是一颗完全对称的理想的骰子。这正如在物理学的研究中忽略摩擦力一样：所得的结果仅仅是一个初步近似。伽里略的考虑对于现实骰子说了些什么？

• 对于现实骰子，三颗骰子能够出现的 216 个可能状态都是接近于等可能的。

• 如果这些状态是等可能的，那么掷出总和为 9 点的机会恰好是 216 分之 25。

• 所以对于现实骰子，掷出总和为 9 点的机会恰好是 216 分之 25 左右。

对于灌过铅的骰子，计算结果会有很大偏离。但是通常的骰子、硬币以及诸如此类的东西都是非常近于公正的——在该意义下所有的结果都是等可能的。当然，你必须在摇动骰子或向上抛硬币时尽一些努力。建立在这些公平的装置上的机会游戏可能相当不公平。(参阅第 17 章)

类似地，在你被要求随机地从一个票盒中抽取一张票时，假设每一张票是等可能地被抽取的。如果票是接近于同样的尺寸、形状和质地，而且盒子是充分摇匀的，那么这是相当合理的近似。

5. 复习题

复习题可能包含前几章的内容。

1. 正确还是错误，并解释之：

(a) 如果一颗骰子掷三次，至少得到一个幺点的机会是 1/6+1/6+1/6=1/2。

(b) 如果一枚硬币抛两次，至少得到一次头像的机会是 50%。

2. 从一副充分洗匀的牌的最上面发出两张牌。你有一次如下的选择：

(i)如果两张牌中至少有一张是Q，你赢1美元。

(ii)如果第一张是Q，你赢1美元。

哪一种选择较有利？或者它们是等价的？解释之。

3. 从一副充分洗匀的牌的最上面发出四张牌。有两种选择：

(i)如果第一张牌是梅花，第二张牌是方块，第三张牌是红心和第四张牌是黑桃，你赢1美元。

(ii)如果四张牌是四种不同的花色，你赢1美元。

哪一种选择较有利？或者它们是相同的？说明之。

4. 一个事件有1/2的机会，另一个事件有1/3的机会。用下面每一对短句中的一个填入空白，构成两个正确的句子。写出这两个句子。

“如果你要求 ___(i)___ 发生的机会，检查一下它们是否 ___(ii)___；如果是的，你可以把机会 ___(iii)___ 。”

(i)两个事件中至少一个，两个事件

(ii)独立的，互不相容

(iii)相加；相乘

5. 一副洗好的牌，把最上面两张牌面朝下地放在桌上。下列结论是否正确，并解释之。

(a)第一张牌是A的机会有13分之1。

(b)第二张牌是A的机会有13分之1。

(c)两张都是A的机会是1/13×1/13。

(d)两张牌中至少有一张是A的机会是13分之2。

6. 掷一对骰子，求两颗骰子出现相同点数的机会。

7. 在Monopoly(垄断)游戏中，一个游戏者掷两颗骰子，计算它们的点数和，并且移动与此和数同样多个的格子。求该游戏者移动11个格子(不多也不少)的机会。

8. 随机放回地从盒子 [1] [2] [2] [3] [3] 中抽取四次。求至少有

一次抽得[2]的机会。

9. 在随机不放回地抽取的条件下，再做习题 8。

10. 从下面给出的两个盒子的每一个中分别随机地抽取一张票：

(A) | [1] [2] [3] |　　　　(B) | [1] [2] [3] [4] |

求以下各事件的机会：

(a)从 A 抽得的票的号码大于从 B 抽得的票的号码。

(b)从 A 抽得的票的号码等于从 B 抽得的票的号码。

(c)从 A 抽得的票的号码小于从 B 抽得的票的号码。

11. 事件 A 的无条件概率为 1/3；事件 B 的无条件概率为 1/10。正确还是错误，解释之：

(a)如果 A 与 B 是独立的，它们一定也是互不相容的。

(b)如果 A 与 B 是互不相容的，它们不可能是独立的。

12. 有两种选择：

(i)一颗骰子掷 60 次，每次如果出现幺点或 6 点，你赢 1 美元；掷出其它点数，你什么也没有赢得。

(ii)随机放回地从盒子 | [1] [1] [0] [0] [0] [0] | 中抽取 60 次。

在每次抽取时，将用美元付给你票上所示的金额。

哪一种选择较好？还是它们是等价的？简短地说明之。

13. 有两种选择：

(i)你抛一枚硬币 100 次；每次抛时，如果出现头像，你赢得 1 美元；如果出现背面，你输掉 1 美元。

(ii)你随机放回地从盒子 | [1] [0] | 中抽取 100 次。每次抽取时，将(用美元)付给你票上的数额。

哪一种选择较好？还是它们是等价的？简短地说明之。

6. 小结

1. 在计算机会时，一种有用的策略是写出机会过程能够出现的全部可能状态的一览表。如果这样做非常困难，至少写出一些典

型的状态，并且算出状态的总数。

2. 假如两件事是互不相容的，这两件事中至少出现一件的机会等于它们各自机会的和。否则，机会相加将会给出一个错误的答案——重复计算。

3. 如果你计算某一个结果的机会有困难，尝试一下计算它的对立结果的机会；然后从100%中减去后者。

15
二项系数

人是一支芦笛，不过是一支会思考的芦笛

——Blaise Pascal（法国，1623－1662）

1. 引言

本章讲述如何回答如下问题：

- 一枚硬币抛四次。恰好得到 1 次头像的机会是多少？
- 一颗骰子掷十次。恰好得到 3 个幺点的机会是多少？
- 一盒装有 1 颗红弹子和 9 颗绿弹子。随机放回地抽取五次。恰好有两次取得红弹子的机会是多少？

这些问题都是类似的，而且能够用 Pascal 和 Newton 发现的二项系数[①]来解。将以弹子为例阐明这一方法。问题是求从盒子中抽取 5 次得到 2 次（不多也不少）红弹子的机会，所以其它 3 次必定是绿的。发生这的一种可能方式是前 2 次抽得红弹子和后 3 次抽得绿弹子，以 R 记红弹子和 G 记绿弹子，这个可能方式可写成

R R G G G

当然，还有别的许多方式出现两次红弹子。例如，第 2 次和第

5 次抽得红弹子，其余各次都是绿弹子：

GRGGR

为求出它的机会，必须找出一切可能的方式，计算每一个的机会，然后用加法规则把这些机会都加起来。第一个任务似乎处理起来有点难度，因此先放一放而转到第二个任务。RRGGG 型的机会是

$$\frac{1}{10}\times\frac{1}{10}\times\frac{9}{10}\times\frac{9}{10}\times\frac{9}{10}=(\frac{1}{10})^2(\frac{9}{10})^3$$

这由乘法规则得出。在每次抽取时，红弹子的机会是 1/10，绿弹子的机会是 9/10。

类似地，GRGGR 型的机会等于

$$\frac{9}{10}\times\frac{1}{10}\times\frac{9}{10}\times\frac{9}{10}\times\frac{1}{10}=(\frac{1}{10})^2(\frac{9}{10})^3$$

GRGGR 型有像 RRGGG 型一样的机会。事实上，每一个具有 2 颗红弹子和 3 颗绿弹子的型式都有相同的机会 $(1/10)^2(9/10)^3$，因为在乘积中 2 颗红弹子提供了 $(1/10)^2$，而 3 颗绿弹子提供了 $(9/10)^3$。由此推出，所有型式的机会之和等于型式的个数乘以共同的机会。

那么有多少型式呢？每一型式是以某种次序写下一行 2 R 和 3 G 所确定的，我们需要求出这样做的不同方式的个数。这个数称做二项系数，并且它有公式：

$$\frac{5\times4\times3\times2\times1}{(2\times1)\times(3\times2\times1)}$$

这个数等于 10。换句话说，2R 和 3G 有 10 种不同的型式。因此，抽得恰好 2 颗红弹子的机会是

$$10\times(\frac{1}{10})^2(\frac{9}{10})^3\approx7\%$$



Isaac Newton(英格兰,1642—1727)

来自 Warder 的收藏;G. Kneller 画,W. T. Fry 按画雕刻。

二项系数看起来比较凌乱。数学家通过引入一种方便的符号来避开这种凌乱。他们在一个数的右端放一个惊叹号(!)去表示这个数与它前面的所有正整数都一起相乘的结果。例如:

$$1! = 1$$

$$2! = 2\times1 = 2$$

$$3! = 3\times2\times1 = 6$$

$$4! = 4\times3\times2\times1 = 24$$

等等。惊叹号读作"阶乘",从而 4! =24 读作"4 阶乘等于 24。"

使用这个符号,二项系数变得较容易写出:

$$\frac{5!}{2!3!}$$

记住这式子表示:把 2 个 R 和 3 个 G 排成一行的不同方式的个数。分子中的 5 是分母中 2 与 3 的和。二项系数总是取这种形式。例如 4 个 R 和 1 个 G 排成一行的方式个数是

$$\frac{5!}{4!1!} = 5$$

这 5 个型式是

RRRRG　RRRGR　RRGRR　RGRRR　GRRRR

把 5 个 R 和 0 个 G 排成一行有多少种方法？只有一种方法：RRRRR。利用这个算法自动地给出

$$\frac{5!}{5!\ 0!}$$

但是我们还没有讲过 0! 的含义是什么。0! =1 是数学上的约定。利用这一约定，该二项系数等于 1。

二项系数和阶乘都会迅速地变得很大。例如，把 10 个 R 和 10 个 G 排成一行的方式数由二项系数得出为

$$\frac{20!}{10!\ 10!}=184\ 756$$

然而，其中有许多可以约去：10! =3 628 800；和 20! $=2\times10^{18}$，或者是 2 后面有 18 个 0。（一万亿是 1 后面有 12 个 0。）

习题 A

1. 求 1 个 R 和 3 个 G 排成一行的不同方式个数。写出所有的型式。
2. 求 2 个 R 和 2 个 G 排成一行的不同方式个数。写出所有的型式。
3. 一盒装有 1 个红球和 5 个绿球。随机放回地从该盒中抽取 4 次。求下列结果的机会：

   (a)一次都没有抽得红球。

   (b)恰好出现一次红球。

   (c)恰好出现二次红球

   (d)恰好出现三次红球

   (e)每次都出现红球

   (f)红球至少出现二次
4. 一颗骰子掷 4 次。求下列结果的机会——

   (a)每次都不出现幺点(一个圆点)

   (b)恰好出现一次幺点

   (c)恰好出现二次幺点
5. 一枚硬币抛 10 次。求恰好得到 5 次头像的机会。并且求得到的头像在 4 次

到 6 次之间(包括两端)的机会。

6. 一枚硬币抛 4 次。恰好得到 1 次头像的机会是多少?

7. 一颗骰子掷 10 次。恰好得到 3 个幺点的机会是多少?

这些习题的答案在第 699—700 页上。

2. 二项公式

第一节的推理总结在二项公式中。假设一个机会过程,像一连串试验一样,分阶段进行:一个例子是一颗骰子掷 10 次,每掷一次骰子可当作一次试验。一个感兴趣的事件,在每次试验时可能发生或者可能不发生:骰子可能出现幺点,也可能不出现幺点。问题是要计算这个事件发生某一确定次数的机会。

> 一个事件在 n 次中恰好发生 k 次的机会由二项公式
>
> $$\frac{n!}{k!\ (n-k)!}p^k(1-p)^{n-k}$$
>
> 给出。公式中 n 是试验次数,k 是事件发生的次数,p 是任何一次特定的试验时该事件发生的概率。假设条件是:
>
> - n 的值必须事先规定。
> - p 的值必须对每次试验都是相同的。
> - 试验必须是相互独立的。

例 1. 一颗骰子掷 10 次。恰好得到 2 个幺点的机会是多少?

解 试验次数是事先规定的。它是 10。所以 n=10。感兴趣的事件是掷出幺点。在每次试验中掷出幺点的概率都相同。是 1/6。所以 p=1/6。试验是相互独立的。因此,可以用二项公式。答案是

$$\frac{10!}{2!\ 8!}(\frac{1}{6})^2(\frac{5}{6})^8\approx 29\%$$

例 2. 掷一颗骰子直到出现 6 点为止。如果能够用二项公式

解,求出现 2 次幺点的机会。如果不能用二项公式,说明为什么不能?

解　试验次数不是事先规定的。它可能是 1,如果骰子一开始就出现 6 点的话。或者,它可能是 2,如果骰子先出现 5,随后出现 6 的话。或者它可能是 3。或者它可能是任意一个数。所以就不能使用二项公式。

例 3. 随机放回地从盒子 [1] [1] [2] [3] [4] [5] 中抽取 10 次。但是,就在抽最后一次之前,不管抽取已经进行得如何,将票 [5] 从盒子中拿掉。正确还是错误:恰好抽到 2 次 [1] 的机会是

$$\frac{10!}{2!\ 8!}\left(\frac{2}{6}\right)^2\left(\frac{4}{6}\right)^8$$

解　在本例中,n 是事先规定的,而且试验是独立的,但是,p 的值在最后一次试验时从 2/6 变为 2/5。所以不能用二项公式,从而陈述是不正确的。

例 4. 随机不放回地从例 3 给出的盒子中抽取 4 次。正确还是错误:恰好抽得 2 张 [1] 的机会是

$$\frac{4!}{2!\ 2!}\left(\frac{2}{6}\right)^2\left(\frac{4}{6}\right)^2$$

解　试验是不独立的,所以不能用二项公式。

技术性注. 要计算例 4 中的机会,取一个恰好有 2 张 [1] 的型式,如 [1] [1] [N] [N],其中 N 的含义"不是 1"。它出现的机会是

$$\frac{2}{6}\times\frac{1}{5}\times\frac{4}{4}\times\frac{3}{3}=\frac{1}{15}$$

出乎意外的是,每一个 2 张 [1] 的型式机会都相同。恰好含两张 [1] 的型式有多少个呢?答案是

$$\frac{4!}{2!\ 2!}=6$$

因此,恰好得到两次 [1] 的机会是

$$6\times\frac{1}{15}=\frac{2}{5}$$

3. 复习题

复习题可能包含前几章的内容。

1. 一颗骰子掷 6 次。恰好得到一次幺点的机会是多少?

2. 一颗骰子掷 10 次。一次都不出现 6 点的机会可以由下列计算之一求出。是哪一个,为什么?

   (i) $(\frac{1}{6})^{10}$ (ii) $1-(\frac{1}{6})^{10}$ (iii) $(\frac{5}{6})^{10}$ (iv) $1-(\frac{5}{6})^{10}$

3. 在有四个孩子的家庭里,女孩多于男孩的比例是多少?你可以设想,孩子性别的确定好像随机放回地从盒子

   | M | F | M = 男性, F = 女性

   中进行抽取[②]。

4. 一盒装有 8 颗红弹子和 3 颗绿弹子。随机不放回地抽取 6 次。正确还是错误:抽得 3 颗绿弹子的机会等于

   $$\frac{6!}{3!\ 3!}(\frac{8}{11})^3(\frac{3}{11})^3$$

   简短地说明之。

5. 某俱乐部有 8 个人,有人编制了一张所有可能的由 2 个成员组成的委员会的名单。另一个人编制了一张所有可能的由 5 个成员组成的委员会的名单。正确还是错误:第二张名单比第一张长一些。简短地说明之。

6. 某俱乐部有 8 个人。有人编制了一张所有可能的由 2 个成员组成的委员会的名单。另一个人编制了一张所有可能的由 6 个成员组成的委员会的名单。正确还是错误:第二张名单比第一张长一些。简短地说明之。

7. 一枚硬币抛 10 次。求在前 5 次中恰好出现 2 次头像,后 5 次中恰好出现 4 次头像的机会。

8. 据称一种维他命补充物能帮助大袋鼠学会通过一座特殊的筑有高墙的迷宫。为了检验这是否正确,把 20 只大袋鼠分成 10 对。

在每一对中，随机地选一只大袋鼠摄取维他命补充物；而另一只喂常规的食物。然后测定大袋鼠学会通过迷宫的时间。10 对中有 7 对，处理过的大袋鼠学会通过迷宫的时间比它的未处理过的配对伙伴要快一些。如果实际上维他命补充物没有效果，这样，每对中各动物都有相等可能成为学得快者，恰巧看到有 7 只或者更多只处理过的动物掌握迷宫比它们未处理过的配对伙伴快的概率是多少？

4. 小结

1. 二项系数给出 n 个对象排成一行的方式数，这 n 个对象中 k 个是互相相同的一类，而(n－k)个是互相相同的另一类(例如，红弹子和蓝弹子)，这公式是

$$\frac{n!}{k!\,(n-k)!}$$

2. 一个事件在 n 次中恰好发生 k 次的机会由二项公式

$$\frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k}$$

给出。公式中 n 是试验次数，k 是事件发生的次数和 p 是事件在任何一次特定试验中出现的概率。它的假设条件是

- n 的值必须事先规定。
- p 必须对每次试验都是同一值。
- 试验必须是独立的。

第五部分　机 会 变 异



16
平均数律

轮盘赌的转盘既无天良又无记性

——Joseph Bertrand(法国数学家 1822—1900)

1. 平均数律讲些什么?

一枚硬币具有 50%的机会出现头像,在抛了许多次以后,头像数应当等于背面数:这不就是平均数律所讲的吗? 南非数学家 John Kerrich 在困难的情况下发现这一点。当第二次世界大战爆发时,他正访问哥本哈根,在他打算飞往英格兰之前两天,德国人入侵丹麦,Kerrich 被拘留在 Jutland 集中营度过战争的余下岁月,为了消磨时间他进行了一系列概率论的试验①,一个试验的内容是抛一枚硬币 10 000 次,得到他的应允,某些结果概括在下面表 1 和图 1 中,Kerrich 的结果对于平均数律讲了些什么呢? 为发现这点,让我们佯装正处在第二次世界大战后期,Kerrich 被邀去向丹麦国王说明平均数律。现在,他与他的助手正在讨论这次邀请。

助手. 这样看来,你是打算把平均数律告诉国王了。

Kerrich. 是的。

助手. 但是有什么可告诉的呢？我的意思是，每一个人都知道平均数律，不是吗？

Kerrich. 好吧。那就请你告诉我平均数律讲些什么。

助手. 行，假设你在抛一枚硬币，如果你得到了大量头像，那么背面就开始走俏，或者如果你得到太多背面，那么头像的机会就上升，从长远来看，头像数与背面数扯平。

Kerrich. 这不正确。

助手. 这不正确，你指的是什么意思？

Kerrich. 我的意思是，你说的都是错的。首先，对于一枚公正的硬币，不管发生什么，头像的机会都保持在50%。在列出的结果中不管有2个头像，还是有20个头像，下一次得到头像的机会仍然是50%。

助手. 我不相信。

Kerrich. 行，譬如说，考虑一串4个头像，我仔细检查了我前2000次抛硬币的情况，硬币连续出现4次的地方有130处；其中69串后面紧跟着出现一个头像，只有61串后面紧跟着出现一个背面，一串头像之后更可能地恰好在下次不出现背面。

助手. 您总是告诉我这些我不相信的事情，您打算告诉国王什么呢？

Kerrich. 好吧，我抛这枚钱币10 000次，我得到大约5000次头像，确切的数字是5067，差额67少于抛的次数的1%，这里在表1中我有记录。

助手. 是的，但是67次头像是一个很大的头像数，倘若那是平均数律能做到的最好的结果，国王将不会留下深刻的印象。

Kerrich. 那么你的见解是什么呢？

表 1 John Kerrich 的抛硬币试验，第一列表示抛的次数，第二列是头像数，第三列表示差：头像数一抛的次数的一半。

抛数	头像数	差	抛数	头像数	差
10	4	−1	600	312	12
20	10	0	700	368	18
30	17	2	800	413	13
40	21	1	900	458	8
50	25	0	1 000	502	2
60	29	−1	2 000	1 013	13
70	32	−3	3 000	1 510	10
80	35	−5	4 000	2 029	29
90	40	−5	5 000	2 533	33
100	44	−6	6 000	3 009	9
200	98	−2	7 000	3 516	16
300	146	−4	8 000	4 034	34
400	199	−1	9 000	4 538	38
500	255	5	10 000	5 067	67

助手. 另外再抛这枚硬币 10 000 次，总共 20 000 次，头像数应该相当接近于期望数，终究，最后头像数与背面数必须相等，对吗？

Kerrich. 这点以前你说过，而它是错的，看一下表 1，在抛 1 000 次中，头像数与期望数之间的差是 2，而抛 2 000 次，差上升到了 13。

助手. 那只是一次意外，到抛了 3 000 次，差只有 10 了。

Kerrich. 那恰好是另一次侥幸，抛 4 000 次，差是 29，抛 5 000 次，差是 33，的确，在抛 6 000 次时，它回落到 9，但是看一下图 1：从抛 1 000 次到 10 000 次机会误差相当稳定地向上爬升，并在临近末尾时直线上升。

助手. 那么平均数律表现在什么地方呢？

Kerrich. 随着抛的次数很大，头像数与期望数之间的差可能按

照绝对值是相当大[②]。但是与抛的次数相比，差可能十分小。那就是平均数律，正如我所说，67 只是 10 000 的很小部分。

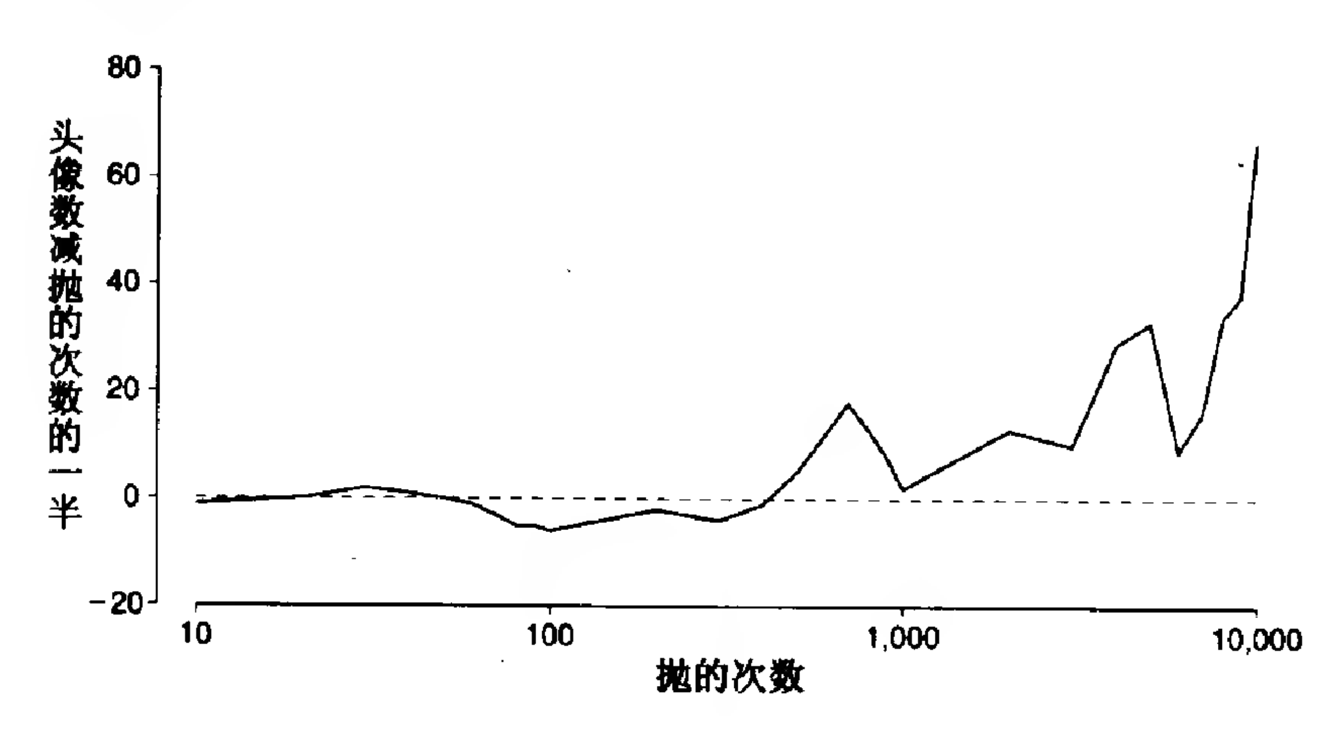
助手. 我不理解。

Kerrich. 瞧，你期望抛 10 000 次可得到 5 000 次头像，对吗？

助手. 对。

Kerrich. 但是不完全如此，你只能期望得到 5 000 次左右头像。我的意思是，你可能正好得到 5 001，或者 4 998，或者5 007。这与 5 000 相差的量是我们所称的“机会误差”。

图 1 Kerrich 的抛硬币实验。“机会误差”头像数－抛的次数的一半相对于抛的次数的图。当抛的次数增加，机会误差趋于上升。水平轴没有按比例绘制。



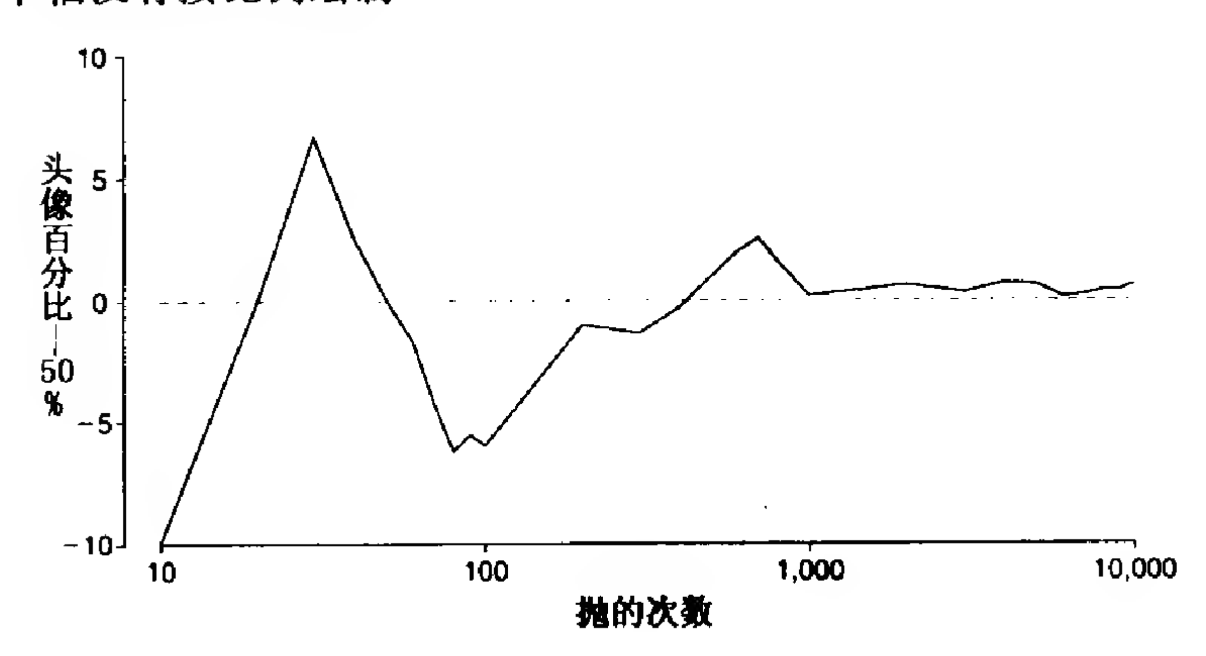
助手. 你能更具体一点吗？

Kerrich. 让我写一个等式：

头像数＝抛的次数的一半＋机会误差

这个误差按照绝对值可能是大的，但是与抛的次数相比是小的，看一下图 2，那就是平均数律，就在那里。

图 2 机会误差表示成抛的次数的百分比，当抛的次数增加时，这个百分比下降，换句话说，机会误差相对于抛的次数变得较小，水平轴没有按比例绘制。



助手。　嗯…；但是如果你抛另外 10 000 次将会发生什么情况呢？那时你将有 20 000 次结果供你计算。

Kerrich.　抛 20 000 次，机会误差可能会更大一些，不过不大可能是 10 000 次时的机会误差的两倍，按绝对值，机会误差将会大一些，但是按抛的次数的百分比，它将会比较小。

助手.　不错，请再告诉我平均数律讲些什么。

Kerrich.　它讲，头像数将是抛的次数的一半左右，允许有一个机会误差。当抛的次数增加时，机会误差变大一些，但是与抛的次数相比，它变小一些。

助手.　您能给我有关机会误差可能有多大的概念吗？

Kerrich.　好，抛 100 次，机会误差的大小可能是 5 左右，抛 10 000 次，机会误差的大小可能是 50 左右，用 100 乘抛的次数时，机会误差的可能量只乘 $\sqrt{100}=10$。

助手.　所以，你说的是，当抛的次数增加时，头像数与抛的次数的一半之间的差变得较大，但是头像的百分比与 50% 之间的差却变得较小。

Kerrich.　确实如此。



习题 A

1. 设计了一台机器可自动抛一枚硬币和记录头像数。抛了 1 000 次，有 550 次头像，请按绝对值与按抛的次数的百分比两种情况下表达机会误差。
2. 习题 1 中的机器，在抛了 1 000 000 次后得出 501 000 个头像，按相同的两种方式表达机会误差。
3. 一枚硬币抛 100 次，出现头像 53 次，但是最后 7 次都抛出头像，正确还是错误：下一次抛出头像的机会略少于 50%，解释之。
4. 抛一枚硬币，如果头像数恰好等于背面数，你赢得 1 美元，哪一种较有利：抛 10 次还是抛 100 次？或它们是相同的？说明之。
5. 抛一枚硬币，如果出现头像的百分比在 40% 与 60% 之间，你赢得 1 美元，哪一种较有利：抛 10 次还是抛 100 次？说明之。
6. 抛一枚硬币，如果有多于 60% 的头像，你赢得 1 美元，哪一种较有利：抛 10 次还是抛 100 次？说明之。
7. 一个内华达轮盘赌的转盘，38 次中有 18 次机会球落入一个红色球囊，转盘将被旋转许多次，有两种选择：

(i)旋转 38 次，如果球落入红色球囊 20 次或 20 次以上，你赢得 1 美元。

(ii)旋转 76 次，如果球落入红色球囊 40 次或 40 次以上，你赢得 1 美元。

哪一种选择较有利？或它们是相同的吗？说明之。

下面三个习题内容是随机地从一个盒子中作抽取，这曾在第 13 章中叙述过，而且将在后面第 3 节中再次进行讨论。

8. 一盒装有 20%红弹子和 80%蓝弹子。随机放回地抽取 1 000 颗弹子，下列陈述之一是正确的：哪一个？为什么？

(i)恰好有 200 颗弹子将是红色的。

(ii)大约有 200 颗弹子将是红色的，允许有一打左右的误差。

9. 重复习题 8，如果抽取是随机不放回地进行，并且盒子装有 50 000 颗弹子。

10. 从下面所示的两个盒子中随机地抽取 100 张票。在每次抽取时，将以美元支付给你票上所示的钱数。（如果抽得负数，将从你那里取走票上所示的钱数。）哪个盒子较有利？或它们是相同的吗？

(i) | −1 | −1 | 1 | 1 |　　　(ii) | −1 | 1 |

11. (难)看图 1，如果 Kerrich 继续抛下去，图会在某个时候取负数吗？

这些习题的答案在第 700—701 页上。

2. 机会过程

Kerrich 的助手曾经努力解决机会变异问题，他逐渐看出，在大数量次地抛一枚硬币时，实际头像数可能不同于期望数，但是他不知道这差异预期有多大，一种计算差异的可能大小的方法将在下一章中给出。当这一方法产生时，它适用于许多不同的场合，例如，在轮盘赌时，它能用来知道庄家应当期望赢多少(第 17 章)；或者它能用来知道一次抽样调查可能的精确度(第 21 章)。

这里共同成分是什么？回答是，所有这些问题都是关于通过机会过程确定的数[3]。在 Kerrich 的试验中取头像数，机会随着每次抛硬币而降临，如果你重复试验，抛出的结果是不同的，因而头像数也是不同的。第二个例子：轮盘赌的输或赢的金额，旋转轮盘是一个机会过程，而且赢或输的金额依赖于结果。再次旋转，赢者又变为输者。最后一个例子：在选民的一个随机样本中民主党人的比

率，一个机会过程被用于抽取样本。因此，样本中民主党人的数目是由抽取的运气所决定的。取另一个样本，你将会得到稍许不同的比率。

这种数受机会影响到什么程度呢？这类问题在统计学中肯定会一次又一次地碰到，下面几章中将提出一个一般对策，两个主要思想：

· 在所研究的过程（选举投票例中的选民抽样）与从一盒中随机抽取数之间寻找类似处；

· 把你要知道的有关（譬如在民主党选民的估计中）的变异与从一只盒中抽得的数之和中的机会变异之间建立联系。一个机会过程与从一盒中作抽取，这两者之间的类似称作盒子模型，要点是从盒子中抽得数之和的机会变异在数学上较容易分析，许多较复杂的过程以后都能够通过这种类似来处理。

3. 抽得数之和

本节的目的是提出两个例子来说明如下的过程：有一盒票，在每张票上写着一个数。然后随机地从该盒中抽取一些票，把这些票上的数加起来，例如，取盒子

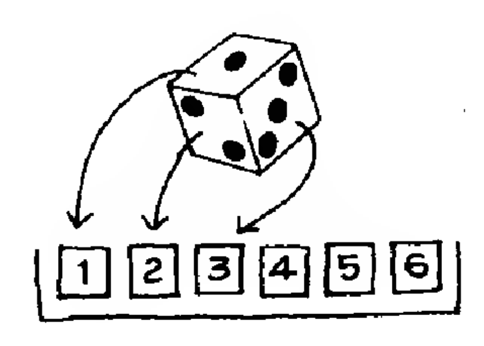
1	2	3	4	5	6

设想随机放回地从该盒中抽取两次：你摇动盒子使得票子混匀，随机地捡起一张票，记下票上的数，把它放回盒中，随后你再次摇动盒子，随机地作第二次抽取，“有放回”一词提醒你在重新抽取之前把票放回盒中，把票放回能使你在同样条件下一次又一次地抽取。（有放回和不放回抽取也曾在第 13 章中讨论过。）

随机放回地抽取了两次，你把这两个数相加起来，例如，第一次抽得的可能是 3 和第二次可能是 5：于是两次抽得数之和为 8。或者第一次抽得的可能是 3 和第二次也是 3：于是两次抽得数之和是 6，存在着许多别的可能性。因此，和是受机会变异影响

的：如果抽取出现一种状态，那么和是一种情况；如果抽取出现不同状态，那么和也是不同的。

首先，这个例子可能觉得有点象是造出来的，但是它恰好象 Monopoly 游戏中的一个回合：你掷一对骰子，把两个骰子出现的点数加起来，并且移动那么多个方格，掷一颗骰子正好象从盒子中捡起一个数。



其次，设想从同一盒子

1　2　3　4　5　6

抽取 25 次，当然，抽取必须有放回地进行，它们之和将大概有多大？找到这个数的一个最直接方式是配制这种盒子和从盒中抽取票。我们编制了计算机程序去做这个试验[4]。第一次抽得 3，第二次抽得 2，第三次抽得 4。它们的全部是：

3 2 4 6 2　3 5 4 4 2　3 6 4 1 2　4 1 5 5 6　2 2 2 5 5

这 25 次抽得数之和是 88。

当然，如果抽得数不同，它们之和也将不同。这样，我们用计算机重复整个过程 10 次，每次随机有放回地从盒中抽取 25 张，计算它们的和，这 10 个结果是：

88　84　80　90　83　78　95　94　80　89

机会变异很容易看出，第一个和是 88，第二个和下降到 84，第三个甚至下降得更多为 80，值的范围从低的 78 到高的 95。

原则上，这和可能小到 $25\times1=25$，大到 $25\times6=150$。但是，事实上 10 个观察值都在 75 到 100 之间。随着更多的重复，这能保持吗？所得和在 75 与 100 之间的机会恰是多少？这类问题将在后面两章中解决。

从一盒中抽得数之和是本节所讨论的如下过程的一种速写：

- 随机地从盒中抽取一些票。
- 把票上的数相加[5]。

习题 B

1. 随机有放回地从盒子 | 1 2 | 中作 100 次抽取，47 次抽得的是 1 其余 53 次是 2，和是多少？
2. 随机有放回地从盒子 | 1 2 | 中抽取 100 次。
   (a)和可能有多小？多大？
   (b)你期望票 1 出现大约多少次？票 2 出现多少次？
   (c)你期望和大约是多少？
3. 随机有放回地从盒子 | 1 2 9 | 中抽取 100 次。
   (a)和可能有多小？多大？
   (b)你期望和大约是多少？
4. 随机放回地从下面三个盒子之一中抽取 100 次，你的事情是猜测和是多少，如果你在偏差为 10 范围内猜对，你赢得 1 美元。对各个盒子，你的猜测是多少？哪一个盒子最有利？哪一个盒子最不利？

   (i) | 1 9 | (ii) | 4 6 | (iii) | 5 5 |
5. 一盒装有 200 张标上"1"的票和 200 张标上"0"的票，从该盒中随机地抽取 100 次，你的任务是猜测抽得数之和，如果你在偏差为 5 范围内猜对，你赢 1 美元，有两种选择：
   (i)抽取是有放回地进行的。
   (ii)抽取是不放回地进行的。
   在每种情况下说出你的猜测。选哪一种较有利？
6. 从下面给出的两个盒子的一个中作 50 次随机有放回的抽取。每次抽取，将用美元付给你票上所示的金额；如果抽到一个负数，从你那里拿走这款额。哪一个盒子较有利？或它们是相同的吗？解释之。

   (i) | −1 2 | (ii) | −1 −1 2 |
7. 你在赌场赌了 4 次。第一次你赢了 4 美元，第二次输了 2 美元，第三次赢了

5 美元，第四次输了 3 美元。下面算式中哪一个告诉你结果赢了多少(正确算式可以不只一个)?

(i) $ 4+ $ 5− $ 2− $ 3

(ii) $ 4+(− $ 2)+ $ 5+(− $ 3)

(iii) $ 4+ $ 2+ $ 5− $ 3

(iv) − $ 4+ $ 2+ $ 5+ $ 3

这些习题答案在第 701 页上。

4. 建立一个盒子模型

本节的目的是构造一些盒子模型为以后用作实践。从盒中抽得数之和是许多统计方法的关键部分，所以你的眼睛要牢牢盯在这个和上，在构造一个盒子模型时，有三个问题要回答：

- 什么数字进入盒中?
- 每一种数字有多少张?
- 抽取多少次?

一个盒子模型的目的是分析机会变异，这种机会变异在任何赌场都能以它的最明显的形式看到。因此，本节的焦点将集中在机会游戏——轮盘赌的盒子模型上。一个 Nevada 轮盘赌的转盘有 38 个球囊；一个标着数 0，另一个标着数 00，而其余球囊标着从 1 到 36 的数。赌桌上收付钱的人旋转转盘并将一只球放到盘里。球等可能地落入 38 个球囊中任意一个。在它落入之前，赌注可以下在(305 页)图 3 所示的表上。

一种打赌是红-和-黑，除了涂了绿色的 0 和 00 外，轮盘赌的转盘上的数字交替地涂了红色和黑色。如果你把 1 美元赌注下在红色上，假定一个红色数字出现，你在收回这一美元的同时还赢得另外一美元。如果出现绿色或黑色数字，赌桌上收付钱的人微笑着，而且迅速用钱耙取走你的钱。

假设你身临拉斯维加斯(Las Vegas)的金块(Golden Nugget)赌场。你恰好把 1 美元放在红色上，赌桌上收付钱的人旋转轮盘。

可能看起来很难计算机会，但是一个盒子模型将会帮助你，那些数进入盒子呢？因为你或者赢得 1 美元，或者输掉 1 美元，所以票上必须标上或者＋＄1，或者－＄1。

第二个问题是，每一类有多少张呢？如果 18 个红色数字中有一个出现，你就赢，如果 18 个黑色数字中有一个出现，你就输。不过，如果出现 0 或 00，你也输。而且这也就是庄家占上风的地方。你赢的机会只是 38 中 18，而输的机会是 38 中 20。所以有 18 张 ＋＄1 和 20 张 －＄1 。盒子是

18 张 ＋＄1	20 张 －＄1

就所涉及到的机会问题来说，把 1 美元赌注下在红色上正好像随机地从盒子中抽取一张票，盒子模型的最大优越性在于，所有不相干的细节——轮盘、赌台和收付赌钱的人的微笑——都已经删去，并且你能看到非常残酷的事实：你有 18 张票，而他们有 20 张票。

那是一次赌博。假如你玩 10 次轮盘赌，每次下一美元赌注于红色，那么可能发生什么情况呢？最终你将会领先或者落后某一金额。这一金额称为你的*净利*。如果你领先，净利是正的；如果你落后，净利是负的。

为了计算机会，必须把净利同盒子联系起来。现在，每次下注，你赢或者输某一金额。这 10 次输赢好像从盒子中作 10 次随机有放回的抽取。（放回票使得每次抽取的机会与转盘一样是相同的。）净利——输或赢的总金额——正好是这 10 次输一赢数的和。因此，10 次赌博中你的净利恰好跟从盒子

18 张票 ＋＄1	20 张票 －＄1

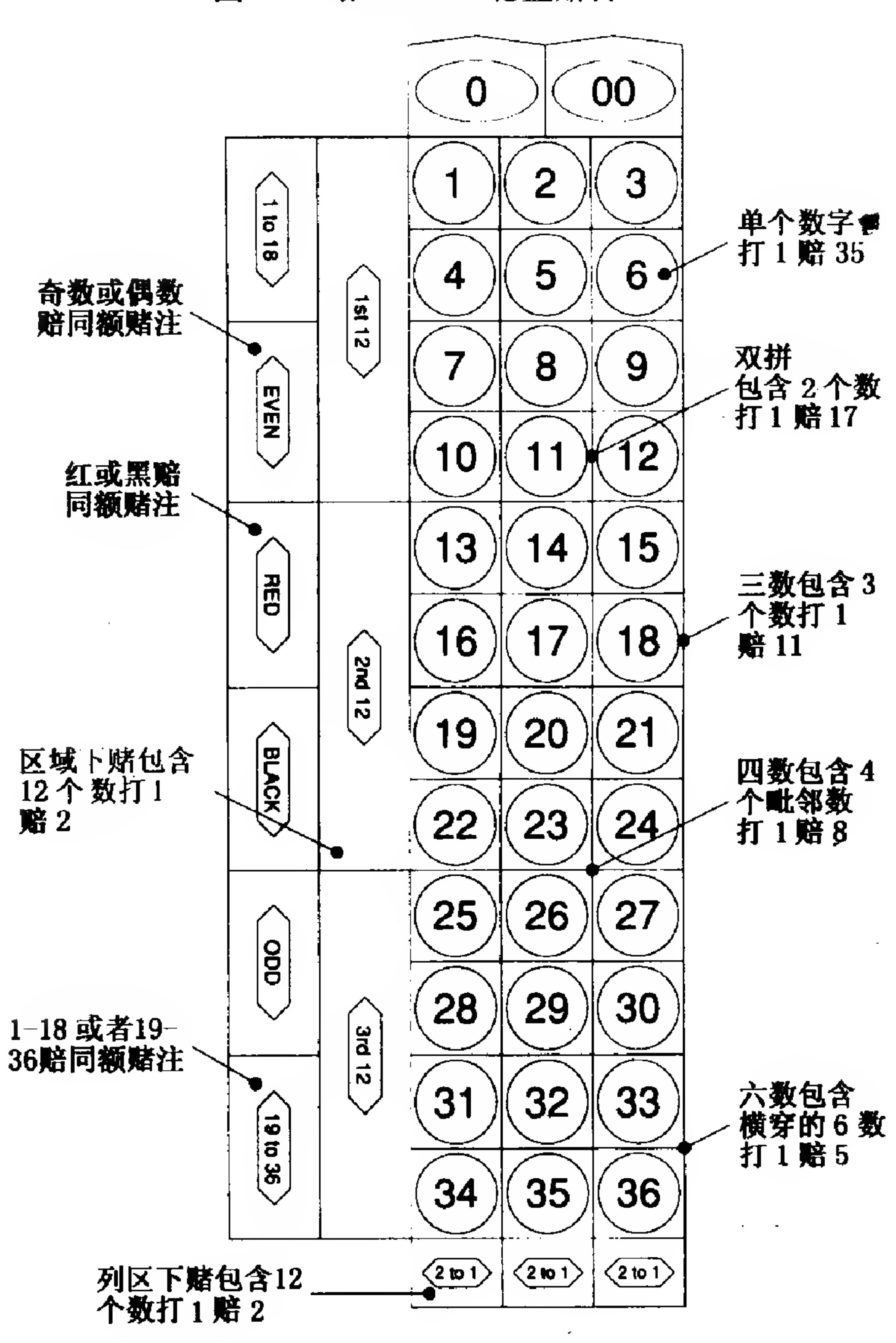
中随机放回地抽取 10 次所得数之和一样。

这是我们的第一个模型，所以更仔细地去考察这一模型是个好主意。例如，假如赌 10 次呈现这样的形式：

R R R B G　　R R B B R

(R 表示红,B 表示黑,而 G 遍示绿——庄家数 0 和 00。)

图 3 一张 Nevada 轮盘赌台



轮盘赌是输掉你的钱的一种愉快、轻松自在和高度舒服的方式。

——Jimmy The Greek(旧金山记事报,1975 年 7 月 25 日)

下面表 2 给出了 10 次相应的输—赢数与净利。

表 2 净利、这是输—赢数的累积和

各次赌博状态	R	R	R	B	G	R	R	B	B	R
输-赢数	+1	+1	+1	−1	−1	+1	+1	−1	−1	+1
净　利	1	2	3	2	1	2	3	2	1	2

顺着净利向前。当你得到红色，输—赢数是+1，且净利上升 1。当你得到黑色或绿色，输—赢数是−1，且净利下降 1。净利恰好是输-赢数之和，而且这与盒子中抽得数一样。因此，净利就好象是从盒子中的抽得数之和。这一局有一个愉快的结果：你赢了 2 美元。如果你继续赌下去。要想知道将会发生什么情况，请阅下一章。

例 1. 如果你把 1 美元赌注下在 Nevada 轮盘赌的一单个数字上，而且这一数字出现，你取回 1 美元赌注同时赢得 35 美元。如果出现任何别的数字，你输掉这块美元，赌徒称这单个数字打 1 赔 35。如果你在轮盘赌上赌 100 次，每次都下 1 美元赌注在数 17 上。你的净利就好象是随机放回地从盒子________作________次抽取所得数之和。填充空白。

解. 什么数进入盒子呢？为回答这个问题，考虑这种赌局的一次下注，你放 1 美元筹码在数 17 上。如果球落入球囊 17，你将得 35 美元。如果它落入任何其它球囊，你将失去 1 美元。因此，盒中必须有票 [$35] 和 [−$1]。

> 盒中的票标明在单次下注中，能够赢或输的各种不同金额。

每一种票有多少张呢？继续考虑单次下注。你只有 38 比 1 的赢的机会；所以抽得 [$35] 的机会必须是 38 次中 1 次，38 次中你有 37 次机会输掉，所以抽得 [−$1] 的机会必须是 38 次中的 37 次。这盒子是

| 1 张票 [$35]　　37 张票 [−$1] |

从盒中抽得任何一个特定数字的机会必须等于单次下注时赢得那一金额的机会(数学上"赢"一个负金额等价于大多数人所称的输这个金额。)

抽取多少次呢?你下注 100 次,所以抽取数必定是 100。在每次抽取后的票必须放回,从而不改变机会。

抽取的次数等于下注参赌的次数

所以在 100 次下注中的净利恰好与从盒子

1 张票 $35	37 张票 −$1

中随机放回地抽取 100 次所得数之和相同。解毕。(期望净利将在 314 页习题 3 中得出。

习题 C

1. 考虑下面三种情况:

   (i) 一盒装有 1 张标有 0 和 9 张标有 1 的票,随机地抽取一张票。如果它显示 1,你赢得一头熊猫。

   (ii) 一盒装有 10 张标有 0 和 90 张标有 1 的票。随机地抽取 1 张票。如果它显示 1,你赢得一头熊猫。

   (iii) 一盒装有 1 张标有 0 和 9 张标有 1 的票。随机放回地作 10 次抽取。如果抽得数之和等于 10。你赢得这头熊猫。

   假设你想要这头熊猫,哪一种较有利——是(i)还是(ii)?或它们是相同的吗?关于(i)与(iii)又是怎样呢?

2. 从盒子

0	1	2	3	4	5	6	7	8	9

   中随机地抽取 1 张票。此票是 0 的机会是多少?它小于或等于 3 的机会是多少?它大于或等于 4 的机会又是多少?

3. 一赌徒打算玩轮盘赌 25 次,每次都放 1 美元在一个双拼上。一个双拼是两个毗邻的数,如在第 305 页的图 3 中的 11 和 12,如果两个数中任意一个出现,赌徒收回 1 美元同时赢得 17 美元。如果这两个数都不出现。他输掉这

1 美元。所以一个双拼是打 1 赔 17，而在 38 次中赢的机会有 2 次，在赌 25 次中此赌徒的净利就好象是从下面三个盒子的一个中作 25 次抽取所得数之和。是哪一个盒子，为什么？

(i) | 0 | 00 | 标号从 1 到 36 的 36 张票 |

(ii) | $17 | $17 | 34 张 −$1 |

(iii) | $17 | $17 | 36 张 −$1 |

4. 在 chuck-a-luck（一种掷三颗骰子的赌博）的一种形式中，三颗骰子从一个盒中滚出。你可以打赌，所有三颗骰子都出现 6，庄家以 1 赔 36，而且打赌者在 216 次中有 1 次机会赢。假设你作这种打赌 10 次，每次都押 1 美元赌注。你的净利就好象是从盒子______中作________次随机有放回抽取所得数之和。填充空白。

5. 一次有 10 个是非题的小测验：一个正确答案得 1 分，一个错误答案不得分，假如一个学生像抛一枚硬币一样随机地回答。这学生的总得分与随机放回地从盒子________中作________次抽取所得数之和相同。填充空白。

这些习题的答案在 701—702 页上。

5. 复习题

1. 抛一枚硬币若干次，如果头像数恰好等于抛的次数的一半，你赢得 1 美元。哪一种较有利，抛 100 次还是 200 次？说明之。

2. 一盒装有 10 000 张票：4 000 张 0 和 6 000 张 1 。从此盒中随机放回地抽取 10 000 次。下面三种说法中哪一种是这个情况的最佳描述。为什么？

(i)1 的次数正好是 6 000。

(ii)1 的次数很可能等于 6 000，但是也存在不等于 6 000 的一些小的机会。

(iii)1 的次数可能不等于 6 000，但是差异相对于 10 000 可能是非常小的。

3. 在随机不放回地从盒中抽取 10 000 次的条件下重复习题 2。

4. 正确还是错误：如果抛一枚硬币 100 次，头像数不可能恰好是

50 次；但是头像的百分率可能正好为 50%，解释之。

5. 正如遗传学理论所指出的一样，在有两个孩子的家庭里，有接近一半对一半的机会两个孩子是同性别的。这里有两种可能性。

   (i)15 对夫妇，每对有两个孩子，在 10 个或更多的家庭里出现两个孩子是同性别的。

   (ii)30 对夫妇，每对有两个孩子，在 20 个或更多的家庭里出现两个孩子是同性别的。

   哪一种可能性更大一些，为什么？

6. 一次有 25 个多重选择题的小测验。每一个问题有五种可能回答，其中一种是正确的。一个正确答案得 4 分，一个错误答案扣 1 分。一学生用随机猜测的方法来回答全部问题。其总得分将与从盒子________中作________次抽取所得数之和相同。填充空白；说明之。

7. 一赌徒在轮盘赌上赌 50 次，每次都把 1 美元赌注下在四个毗邻数(例如图 3 中 23、24、26、27)上，如果这四个数中有一个出现，她取回这 1 美元赌注和赢得 8 美元。如果出现任何其它数，她输掉这 1 美元。因此，这种赌博是打 1 赔 8，且在 38 次中有 4 次赢的机会。赌 50 次中她的净利与从盒子________中作________次抽取所得之和相同。填充空白；说明之。

8. 一赌徒在轮盘赌上连续输了 10 次。他决定继续赌下去，因为根据平均数律他应当赢一次。一位旁观者劝他放弃，因为当时他的运气欠佳，谁正确？或他们俩都是错误的？

9. 随机放回地从盒子 | −3 | −2 | −1 | 0 | 1 | 2 | 3 | 中抽取 200 次。

   (a)如果 200 个抽得数之和是 30，它们的平均数是多少？

   (b)如果 200 个抽得数之和是 −20，它们的平均数是多少？

   (c)一般地，当告诉你抽得数之和时，你怎样才能算出 200 次抽取所得数的平均数？

   (d)有两种选择：

(i)如果 200 个抽得数之和在－5 与＋5 之间，赢得 1 美元。

(ii)如果 200 个抽得数的平均数在－0.025 与＋0.025 之间，赢得 1 美元。

哪一种较有利？或它们是相同的吗？解释之。

6. 小结

1. 头像数中存在机会误差：

头像数＝抛的次数之一半＋机会误差。

误差按绝对值可能是大的，但是相对于抛的次数是小的。这就是平均数律。



由 Dana Fradon 绘制：1976 纽约人杂志，附图

2. 平均数律能以百分率的措词重述。在抛大数量次时，尽管头像的百分率不可能恰好等于 50％，但是它可能接近于 50％。

3. 补偿对平均数律不起作用。例如，在抛一枚硬币时，一串头像之后出现头像与出现背面是完全一样可能的。

4. 一个复杂的产生一个数的机会过程常常可以模型化为从盒子中作抽取。抽得数之和是一个关键组成部分。

5. 在构造一个盒子模型时要求回答三个基本问题：

- 哪些数字进入盒子？
- 每一类数有多少张？
- 抽取多少次？

6. 对于下同一赌注多次的赌博问题，一个盒子模型能建立如下：

- 盒中的票标明每次下注能赢(+)或输(−)的金额。
- 从盒中抽得任一特定值的机会等于在单次下注中赢得该金额的机会。
- 抽取次数等于下注次数。

于是，净利与从盒中抽取所得数之和相同。

17 期望值与标准误差

如果你相信奇迹，就到 Keno[1] 赌厅去。
——Jimmy，The Greek

1. 期望值

正进行着一个机会过程。它给出一个数，随后另一个数，再另一个数，你几乎将被淹没在随机输出之中，但是数学家在这种紊乱中发现了一些规律：这过程给出的数围绕着期望值变化，其偏离量的大小与标准差相似。如果这过程只进行一次，它所产生的数应当在期望值周围的某处，大约相差一个标准误差左右。

更具体一些，设想通过下列机会过程产生一个数：在一枚钱币抛 100 次中计数头像数，你可能得到 57 次头像。这在期望值 50 之上 7，所以机会误差是 +7，如果你抛另外 100 次，你得到不同的头像数，也许是 46：机会误差将是 −4，第三次重复抛 100 次，可能产生又一个别的数。譬如说是 47：机会误差将是 −3，你的数将偏离

① Keno 克诺是一种纸牌赌博——译者注。

50，一个在大小上与标准误差类似的随机量，它是 5（参阅后面第 5 节）。

期望值与标准误差的公式依赖于产生数的机会过程。本章处理从盒中抽取所得数之和，关于期望值的公式将用一个例子来介绍：随机放回地从盒子

1	1	1	5

中抽取 100 次所得数之和。

此和大约应当多大？为回答这个问题，考虑抽取应该如何进行。盒中有四张票，所以 5 应当在大约$\frac{1}{4}$的抽取次数中出现，而 1 则是$\frac{3}{4}$左右。抽取 100 次，你能够期待得到大约 25 次 5 和 75 次 1，所以抽得之和应当在

$$25\times5+75\times1=200$$

左右。这就是期望值。

期望值的公式是一条捷径。它有两个组成部分：

- 抽取次数；
- 盒中数的平均数，简称“盒平均”。

随机放回地从盒中抽取所得数之和的期望值等于（抽取次数）×（盒平均）

为了解公式后面的逻辑原理，回到例子。此盒的平均数是

$$\frac{1+1+1+5}{4}=2$$

平均说来，每次抽取对于和大约加上 2。抽取 100 次，和必定是 100×2=200 左右。

例 1. 假设你打算去 Las Vegas（拉斯维加）玩 Keno，你特别喜爱把 1 美元赌注下在单个数字上。在你赢时，他们给还你 1 美元，并且再给你 2 美元。在你输时，他们就留下这 1 美元。在 4 次中有 1 次机会你赢①。如果你每次都这样下赌注，在赌 100 次中你应该

大约期望赢(或输)多少?

解. 第一步是写出一个盒子模型。在每次下注时,你要末有$\frac{1}{4}$的机会,净利上升 2 美元;要末有$\frac{3}{4}$的机会,净利下降 1 美元。所以在 100 次下注后你的净利与随机放回地从盒子

\$2	−\$1	−\$1	−\$1

中抽取 100 次所得数之和相同。

此盒的平均数是

$$\frac{\$2-\$1-\$1-\$1}{4}=-\$0.25$$

平均而言,每次赌博使你丧失一个夸脱(25 美分)。赌 100 次,你可能期望输掉大约 25 美元。这就是答案。如果你继续赌下去,赌 1000 次,你应该期望输掉大约 250 美元。你赌的次数愈多,你输得愈多。或许你应该寻找另一种赌博游戏。

习题 A

1. 分别求出随机放回地从盒子

   (a) | 0 | 1 | 1 | 6 | (b) | −2 | −1 | 0 | 2 |

   (c) | −2 | −1 | 3 | (d) | 0 | 1 | 1 |

   中抽取 100 次所得数之和的期望值。
2. 求出 Monopoly 游戏(参阅第 301 页)中,第一次玩时移动格子的期望数。
3. 某人打算在轮盘赌上赌 100 次,每次都把 1 美元赌注下在 17 上。求净利的期望值。(阅第 306 页例 1)
4. 你打算在轮盘赌上赌 100 次,每次都把赌注 1 美元下在红-与-黑上。求你的净利的期望值。(这种赌博偿付与赌注相同的钱,并且你在 38 次中有 18 次机会得胜。阅第 305 页上表 3)
5. 在赌 1 000 次的情况重复习题 4。
6. 一种赌博,如果净利的期望值等于 0,是公平的,结果是平均来说赌博者既不赢,也不输。如果一赌徒把 1 美元的赌注下在轮盘赌的红-与-黑上,而且

得胜，一个慷慨的赌场将会给比 1 美元多一点的钱，他们应当付多少才能使这种赌博是公平的？（提示：令 x 表示他们应当付的 。盒子中装有 18 张票 [x] 和 20 张票 [－＄1]。用 x 写出有关期望值的公式，并使此式等于 0。）

7. 如果一个冒险家在 Royal Oak 赌博时，把 1 磅赌注下在一点上，而且得胜了，球的主人应该付给他多少钱才使赌博是公平的。（规则说明在第 276 页的习题 4 中。）

这些习题的答案在 702 页上。

2. 标准误差

假设从盒子

0	2	3	4	6

中随机放回地抽取 25 次。（盒中的数字无特殊含义；选择它们是为了后面的计算结果为整数。）五张票中每一张应该在抽取次数的大约 $\frac{1}{5}$ 中出现，就是说 5 次中 1 次，所以和应当是

$$5\times0+5\times2+5\times3+5\times4+5\times6=75$$

左右，这就是和的期望值。当然，正如 Kerrich 没有得到恰好为抛的次数一半的头像一样，每一张票也不会恰好出现抽取次数的 $\frac{1}{5}$，该和与期望值将偏离一个机会误差：

和＝期望值＋机会误差

机会误差是高（＋）于或低（－）于期望值的量。例如，如果出现的和为 70，则机会误差为－5。

机会误差可能有多大？答案由标准误差，简写为 SE，给出。

> 一个和可能在它的期望值附近，但是偏离一个其大小与标准误差相似的机会误差

有一个公式可用来计算随机放回地从一盒中作抽取所得数之和的 SE。它叫做平方根法则，因为它含有抽取次数的平方根。在本书余下部分中，提出的所有统计方法都与这个公式有关。

在从装有标上数的票子的盒中作随机有放回地抽取时，抽得数之和的标准误差是

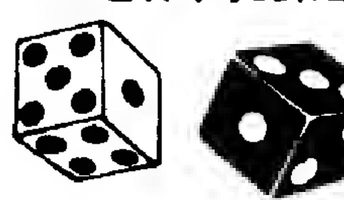
$$\sqrt{\text{抽取次数}} \times (\text{盒子的 SD})$$

公式有两个组成部分：抽取次数的平方根和盒中一系列数的SD，简称“盒子的 SD”。SD 度量盒子中所有数之间的散布程度。如果盒子中存在很大的散布程度，SD 就大，从而很难预测抽取将会出现的结果。因此标准误差也必定大。现在说一下抽取次数。两次抽取所得数之和比单独一次抽取变化多；100 次抽取所得数之和的变化就更大。每次抽取都给和增添了额外的变异，因为你不知道它将会出现什么。当抽取的次数增加时，和将会更难预测，机会误差会更大，从而标准误差也会更大。但是标准误差由于一个乘积因子等于抽取次数的平方根而上升得较慢。例如，100 次抽取所得数之和仅是单独一次抽取的变异的$\sqrt{100}=10$倍。

SD 与 SE 是不同的[②]。SD 适用于一系列数字的散布程度。它可以用第 80 页上所阐述的方法计算。对比之下，SE 适用于机会变异——例如抽取所得数之和中的机会变异。

SD 适合于一列数 **SE 适合于机会过程**

1 2 3 4 5 6



在本节的开头，我们考虑随机放回地从盒子

0	2	3	4	6

中抽取 25 次所得数之和。此和的期望值是 75：和将在 75 附近，相差一个机会误差。这一机会误差可能有多大呢？为求出它，计算标准误差，盒中数的平均数是 3。与平均数的偏差是

$$-3 \quad -1 \quad 0 \quad 1 \quad 3$$

所以盒子的 SD 是

$$\sqrt{\frac{(-3)^2+(-1)^2+0^2+1^2+3^2}{5}}=\sqrt{\frac{9+1+0+1+9}{5}}$$

$$=\sqrt{\frac{20}{5}}=2$$

这度量了盒中的变异。根据平方根法则，由于因子$\sqrt{25}=5$，25 次抽取所得数之和的变化要大些，所以 25 次抽取所得数之和的 SE 是 5×2=10。换句话说，机会误差的可能大小是 10。而且抽得数之和应当在 75 附近，加减 10 左右。一般地，和可能在它的期望值附近，相差一个标准误差左右。

为了经验地说明其含义，我们用一台编好程序的计算机，随机放回地从盒子

0	2	3	4	6

中抽取 25 次。得到

00440　43262　20262　64263　03640

这 25 次抽得数之和是 71。它低于期望值的量为 4，所以机会误差是−4。计算机抽取另外 25 次，并且取它们的和，得 76。机会误差是+1。第三个和是 86，有机会误差 11。实际上，我们用计算机产生了 100 个和，列在表 1 中。这些数都在期望值 75 附近。由于机会误差，它们偏离的大小相似于标准误差 10。

表 1 中的和显著地表明在期望值附近少量的散布，原则上，它们可以小到 0。或者大到 25×6=150。但是，除了一个外其余全部都在 50 与 100 之间，就是说，在期望值的 2.5 个 SE 之内。离期望值的差大于 2 个或 3 个 SE 的观察值很罕见。而且由于公式中的平方根，和的 SE 随着抽取次数的增大而增长得十分缓慢。

平方根法则的基本原理是相互抵销。在例中盒的平均数是 3，所以 0 和 2 低于平均数，4 和 6 高于平均数。每当高于平均数的值发生低于平均数的值也差不多发生，其误差彼此抵销到某种程度。结果，抽得数之和保持比较接近于它的期望值。平方根法则告诉你

可望抵销多少?

表 1　随机放回地从盒子 | 0 2 3 4 6 | 中抽取 25 次所得数之和的 100 个观察值的计算机模拟

重复号	和	重复号	和	重复号	和	重复号	和	重复号	和
1	71	21	80	41	64	61	64	81	60
2	76	22	77	42	65	62	70	82	67
3	86	23	70	43	88	63	65	83	82
4	78	24	71	44	77	64	78	84	85
5	88	25	79	45	82	65	64	85	77
6	67	26	56	46	73	66	77	86	79
7	76	27	79	47	92	67	81	87	82
8	59	28	65	48	75	68	72	88	88
9	59	29	72	49	57	69	66	89	76
10	75	30	73	50	68	70	74	90	75
11	76	31	78	51	80	71	70	91	77
12	66	32	75	52	70	72	76	92	66
13	65	33	89	53	90	73	80	93	69
14	84	34	77	54	76	74	70	94	86
15	58	35	81	55	77	75	56	95	81
16	60	36	68	56	65	76	49	96	90
17	79	37	70	57	67	77	60	97	74
18	78	38	86	58	60	78	98	98	72
19	66	39	70	59	74	79	81	99	57
20	71	40	71	60	83	80	72	100	62

习题 B

1. 随机放回地从盒子

| 1 2 3 4 5 6 7 |

中抽取 100 次。

(a)求抽得数之和的期望值与标准误差。

(b)假设你必须猜测该和是什么,你将猜多少? 你期望偏离约 2、4 还是 20 左右?

2. 你就抛一枚硬币打赌 100 次。如果出现头像,你赢 1 美元。如果出现背面,你输 1 美元。你的净利大约是________,有________左右的误差,利用供选择的数据

−$10　−$5　$0　+$5　+$10

填充空白。

3. 和的期望值是50，其SE为5。将产生和的机会过程重复10次，下面哪一个是它的观察值序列，为什么？

(i)51,57,48,52,57,61,58,41,53,48

(ii)51,49,50,52,48,47,53,50,49,47

(iii)45,50,55,45,50,55,45,50,55,45

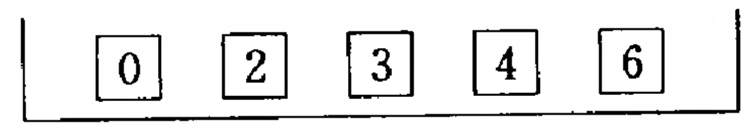
4. 如果计算机继续进行，你认为最终会产生出一个偏离期望值大于3SE的和吗？说明之。

这些习题的答案在702—703页上。

3. 使用正态曲线

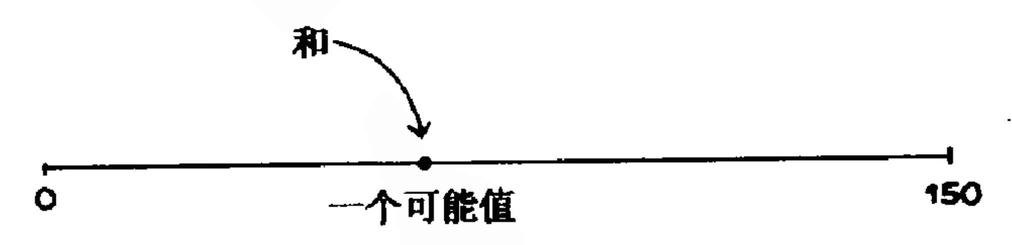
随机放回地从一盒中作大数次抽取，抽得数之和在一个给定范围内的机会是多少？数学家们在试解这类问题时发现了正态曲线。曲线背后的逻辑原理将在下一章中讨论；本节的目的仅在于概叙这一方法，只要抽取次数适当大这方法都适用。基本上是（用期望值和标准误差）先转换成标准单位，随后计算曲线下的面积，就象在第5章中那样。

现在举一例子，假设计算机编好程序从魔盒



中随机放回地抽取25次，并求其和。一次次重复这一过程，并打印出结果。大概有百分之多少的观察值应当在50与100之间？

每一个和将在水平轴上0与25×6=150之间某个地方。



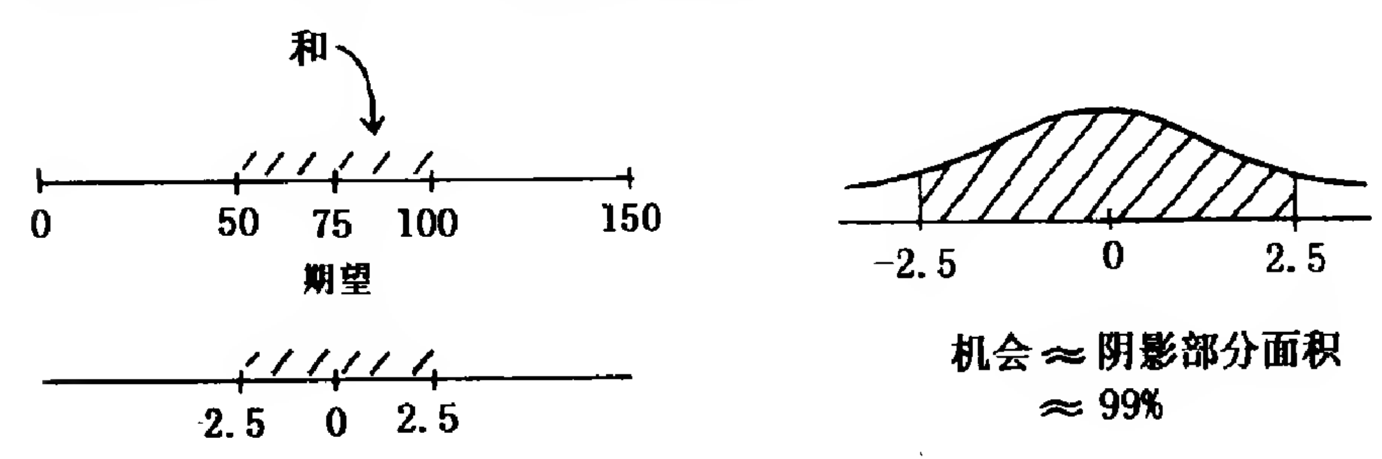
问题是求和将出现在50与100之间的机会。

为求机会，转换成标准单位，并利用正态曲线。就此而论，标准

和

0 50 100 150

单位说出一个数偏离期望值多少个 SE。在本例中,100 在标准单位下变成 2.5。理由是:和的期望值是 75,且 SE 是 10,所以 100 超过期望值 2.5 个 SE。类似地,50 变成—2.5。



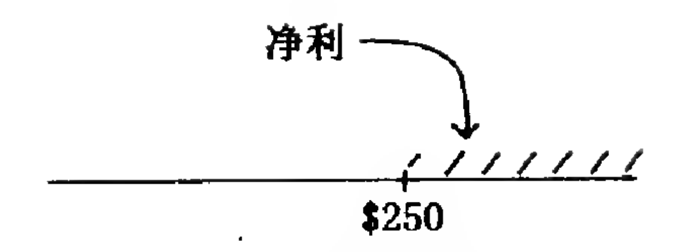
从 50 到 100 之间的区间是期望值的 2.5SE 内的区间,因此和应该大概有 99%的时刻落在那里面。

这就完成了计算。现在讨论某些数据。前面表 1 报导了 100 个和的观察值:它们中大约 99 个应该在 50 到 100 的区间内,事实上它们中确有 99 个是在这区间内。为了取含更少极端的某些范围,大约有 68%的观察值应该在 75－10 到 75＋10 的区间内。事实上,有 73 个观察值在这个区间内,最后,表 1 中大约 95%的观察值应该在 75±20 的范围内,事实上它们中有 98 个在这范围内。(范围包含了端点。)所以这理论看上去颇佳。

例 2. 一个月中,在某一赌场的轮盘赌转盘上独立地赌了 10 000 次,为简单起见,假设赌徒每次赌时都把 1 美元赌注下在红色上,估计庄家从这么多次赌博中所赢多于 250 美元的机会。(红-与-黑付相等的金额,而且庄家在 38 次中有 20 次赢的机会。)

解. 问题是求庄家的净利超过 250 美元的机会。

这里,问题没有肯定盒子模型。要做的第一件事是写出盒子模型。哪些数进入盒子呢? 在每次赌博时,庄家或者赢 1 美元,或者输 1 美元,所以票是 $1 和 －$1 。每一类票有多少张呢? 在每次



赌博时，庄家在 38 次中 20 次赢的机会和 18 次输的机会。因此，这盒子是

20 张票 +$1	18 张票 −$1

赌徒与庄家分属这盒子的对立两边：对庄家有 20 张票有利，而对赌徒 18 张票有利。（参见第 16 章从赌徒的立场对这个模型的讨论。）

应该抽取多少次呢？有 10 000 次赌博，所以必须抽取 10 000 次，庄家的净利与随机放回地从盒中抽取 10 000 次所得数之和相同。这样建立了模型，且使我们处于能够使用平方根法则的地位。

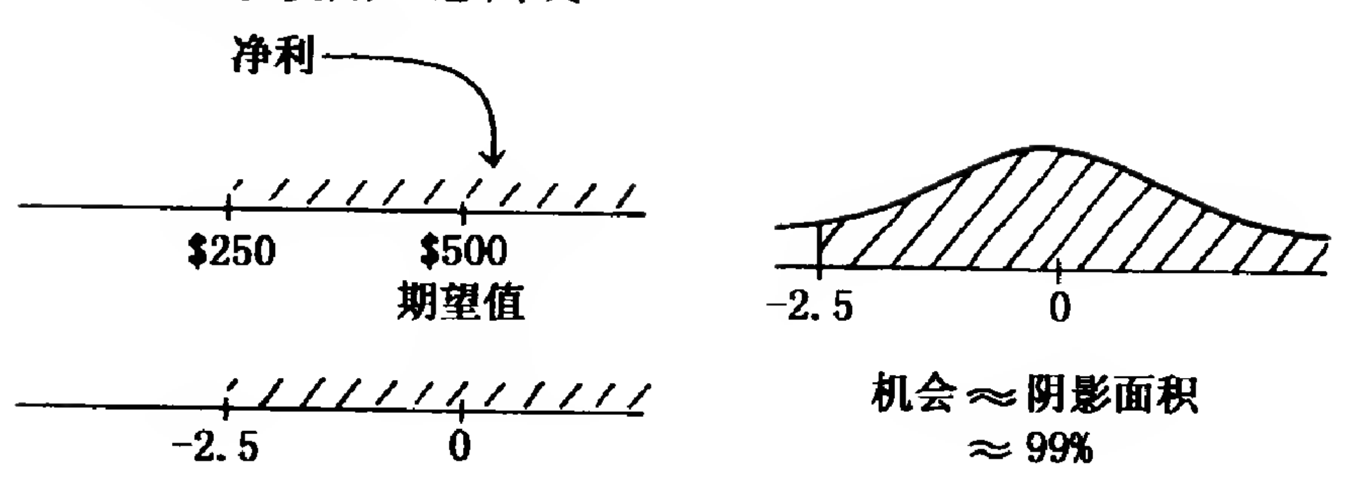
下一步是求净利的期望值。盒中数的平均数是需要的。它等于盒中数的总和除以 38。总之，20 个正数给总和贡献了 20 美元，但是 18 个负数取走了 18 美元。因此，盒的平均数是

$$\frac{20\text{ 美元}-18\text{ 美元}}{38}=\frac{2\text{ 美元}}{38}\approx 0.05\text{ 美元}$$

平均说来，每次抽取给和增加了大约 0.05 美元。因此，10 000 次抽得数之和有期望值 10 000×0.05 美元=500 美元。这说明庄家平均每次赌博可赢大约五美分，因此，在 10 000 次赌博中能够期望赢得大约 500 美元。

接下去是求净利的 SE。这需要盒中数的 SD。因为盒平均数接近于 0，所以盒中每一个数与平均数的偏差恰好都在 1 美元左右。从而盒子的 SD 大约是 1 美元。这 1 美元度量了盒子中的变异。根据平方根法则，10 000 次抽取所得数之和，会变化 $\sqrt{10\ 000}=100$，多倍于 1 美元。10 000 次抽得数之和的 SE 是 100×1 美元=100 美元。庄家能够期望赢得大约 500 美元，加减 100 美元左右的误差。

现在能够使用正态曲线。



解毕。关键的步骤是指明，净利与从盒中抽取所得数之和相同；这提供了平方根法则的逻辑基础。

庄家有大约 99%的机会赢得多于 250 美元。这可能看起来好像不多，但是你必须记住，庄家拥有许多转盘，而且往往有一大群人在每个转盘的每次旋转时进行赌博，下的赌注很多是超过一美元的。庄家能够期望赢得大约 5%放在赌台另一面的钱，而且平方根法则几乎消除了风险。例如，假定庄家开设 25 台赌盘。保守一点，假定每一赌盘在例 2 条件下运行。在这些假定条件下，赌场的期望赢钱数按满载因子 25 上升到 25×500 美元＝12 500 美元。但是它们的标准误差按因子$\sqrt{25}=5$只上升到 500 美元。现在赌场能够几乎肯定——99%的把握——至少赢 11 000 美元。对赌场来说，轮盘赌是一家像杂货店一样安全经营的大商店。

习题 C

1. 随机放回地从盒 | 1 1 2 2 2 4 | 中抽取 100 次。

 (a)能够出现的最小的和可能是________，最大的和是________。

 (b)抽得数之和将在________周围，有________左右的误差。

 (c)和大于 250 的机会差不多是________%。

2. 随机放回地从盒子 | 1 3 3 9 | 中抽取 100 次。

 (a)和可能有多大？有多小？

(b)和在 370 到 430 范围内有多少可能?

3. 你可以随机放回地从盒子 | -1 1 | 中或者抽取 10 次或者 100 次,在下列每一种情况你应当抽取多少次?

(a)当和是 5 或者大于 5 时你赢得 1 美元,其它情况不赢。

(b)当和是 -5 或者小于 -5 时你赢得 1 美元,其它情况不赢。

(c)当和是在 -5 与 5 之间时你赢得 1 美元,其它情况不赢。

不必计算,但是要说明你的理由。

4. 有两种选择:

(i)随机放回地从盒子 | 1 1 5 7 8 8 | 中抽取 100 次。

(ii)随机放回地从盒子 | 4 7 11 13 15 | 中抽取 50 次。

如果支付赢利是

(a)1 美元当和大于或等于 550 时,其它情况不支付。

(b)1 美元当和小于或等于 450 时,其它情况不支付。

(c)1 美元当和在 450 与 550 之间时,其它情况不支付。

哪一种选择较有利?

5. 假设某赌场的轮盘赌一星期内有 25 000 次独立赌博。每次赌时赌徒都把 1 美元赌注下在红色上。赌场从这 25 000 次赌博中赢得多于 1 000 美元的机会最接近于:5%、50%、95%中哪一个?简短地说明之。

6. 假设某人在一次赌博中把 25 000 美元赌注下在转盘的红-与-黑上,从这次赌博中赌场赢得多于 1 000 美元的机会最接近于:5%、50%、95%中哪一个?简短地说明之。

7. 一赌徒在轮盘赌上赌一次,在每个数(包括 0 和 00)上都下 1 美元赌注,这样此人总共下了赌注 38 美元。将会发生什么状况?简短地说明之。

8. 一盒装有六张编号从 1 到 6 的票;此盒的平均数为 3.5,SD 为 1.7,编号为 1、2、3 的票是白色的,而编号为 4、5、6 的票是黑色的。

| 1 2 3 4 5 6 |

某人用下述过程从此盒中选取 100 个数。他拣出白票与黑票分放入各别的盒子中。随后,他随机放回地取 50 张白票。最后,他随机放回地取 50 张黑票。正确还是错误:100 次抽得数之和的标准误差是 $\sqrt{100}\times 1.7$。简短地说明之。

这些习题的答案在 703 页上。

4. 一条捷径

求 SD 可能是伤脑筋的事，但是对于只有两类票的盒子有一条捷径。

> 当盒子中票只表示两个不同数字时，盒子的 SD 等于
>
> $$(\text{较大数}-\text{较小数})\times\sqrt{\begin{matrix}\text{较大数票}\\\text{占的比率}\end{matrix}\times\begin{matrix}\text{较小数票}\\\text{占的比率}\end{matrix}}$$

例如，取盒子 [1] [1] [1] [5]。因为票上只指明两个不同数字 1 和 5，这条捷径可以使用。所以 SD 是

$$(5-1)\times\sqrt{\frac{1}{4}\times\frac{3}{4}}\approx1.73$$

这条捷径所包含的计算比求离平均数的偏差的均方根(参阅第 80 页)要少得多，而且正确地给出同一答案。

例 3. 一赌徒轮盘赌上赌 100 次，每次都把 1 美元赌注下在数 10 上，打 1 赔 35，并且赌徒在 38 次中有 1 次赢的机会。填充空白：赌徒期望赢________美元，有大约________美元的误差。

解. 第一件要做的事是对净利建立一个盒子模型。(参阅第 306 页上例 1。)赌徒的净利是与随机放回地从盒子

1 张 [$ 35]	37 张 [- $ 1]

中抽取 100 次所得数之和相同。

所期望的净利是多少? 这是盒子平均数的 100 倍。盒中数的平均数是它们的总和除以 38。1 张赢的票对总和贡献 35 美元，同时 37 张输的票总共要减去 37 美元。因此平均数是

$$\frac{35\text{ 美元}-37\text{ 美元}}{38}=\frac{-2\text{ 美元}}{38}\approx-0.05\text{ 美元}$$

平均说来，赌徒每次期望输掉大约 5 美分左右，在 100 次赌博中期

望净利为

$$100\times(-0.05\text{ 美元})=-5\text{ 美元}$$

换句话说，赌徒在 100 次赌博中期望输掉大约 5 美元左右。

下一步是求出抽得数之和的 SE：这是盒中 SD 的 $\sqrt{100}$ 倍。可以用所说的捷径，盒中的 SD 等于

$$[35\text{ 美元}-(-1\text{ 美元})]\times\sqrt{\frac{1}{38}\times\frac{37}{38}}\approx 36\text{ 美元}\times 0.16\approx 5.76\text{ 美元}$$

抽取 100 次所得数之和的 SE 是 $\sqrt{100}\times 5.76$ 美元 ≈ 58 美元

赌徒期望输掉大约 5 美元，允许有大约 58 美元的误差。解毕。大 SE 给予赌徒赢许多钱的合乎情理的机会，而这是一种诱惑力。当然，平均说来赌徒是输的；而且大 SE 也意味着赌徒可能会输得很多。

习题 D

1. 对下面每一数列，说出公式是否给出 SD，并说明之。

数列	公式
(a)7、7、7、−2、−2	$5\times\sqrt{\frac{3}{5}\times\frac{2}{5}}$
(b)0、0、0、0、5	$5\times\sqrt{\frac{1}{5}\times\frac{4}{5}}$
(c)0、0、1	$\sqrt{\frac{2}{3}\times\frac{1}{3}}$
(d)2、2、3、4、4、4	$2\times\sqrt{\frac{1}{6}\times\frac{2}{6}\times\frac{3}{6}}$

2. 在第 320-322 页例 2 中，当一赌徒下一美元赌注在轮盘赌的红色时，有一个赌场赢钱的盒子模型。盒子的 SD 为 1 美元左右。求出 SD 到第四位小数。

3. 假设一个赌徒在 Keno 赌博游戏中把 1 美元赌注下在单个数字上（第 313 页例 1）。在 100 次赌博中，赌徒的净利将是________美元，加减______美

元左右的误差。

4. 在 Nevada 轮盘赌台上，"庄家专号"是把赌注下在数 0、00、1、2、3 上。打 1 赔 6，在 38 次中有 5 次赢的机会。

(a)对于 Nevada 轮盘赌台上所有其它的下注，庄家期望从放在台上的每 1 美元中获取 5 美分。问在庄家专号上每 1 美元可期望获取多少？

(b)某人在轮盘赌上赌 100 次，每次都把 1 美元赌注下在庄家专号上。估计此人最终赢钱的机会。

5. 一赌徒在轮盘赌上赌 100 次。有两种可能性：

(i)每次都把 1 美元赌注下在区域上(参阅第 305 页图 3。)

(ii)每次都把 1 美元赌注下在红色上。

区域下注是打 1 赔 2，并且在 38 次中有 12 次赢的机会。赌红色是赔相等金额，而且在 38 次中有 18 次赢的机会。正确还是错误，并说明之：

(a)(i)与(ii)中结果赢钱机会是相等的。

(b)赢 10 美元以上的机会，(i)较大些。

(c)输掉 10 美元以上的机会，(i)较大些。

这些习题的答案在 704 页上。

5. 分组与计数

某些机会过程包含计数，平方根法则能够用来获得计数的标准误差，但是必须正确地建立盒子模型。下面的例子将指出如何做这件事。

例 4. 一颗骰子掷 60 次。

(a)总点数应该是____左右，加减大约____的误差。

(b)6 点出现的个数应该是____左右，加减大约____的误差。

作为说明，表 2 给出一颗骰子抛 60 次的结果；第 1 次抛出 4，第 2 次抛出 5，等等。

表 2　一颗骰子抛 60 次的结果

4 5 5 2 4	5 3 2 6 3	5 4 6 2 6	4 4 2 5 6
1 5 3 1 2	2 1 2 5 3	3 6 6 1 1	5 1 6 1 2
4 4 2 1 4	4 5 2 6 3	2 4 6 1 6	4 6 1 5 2

解. 部分(a)是熟悉的。它涉及做加法。每抛一次都贡献一些点

数，我们把这些数加起来。抛 60 次骰子所得的总点数恰好与从盒子

[1] [2] [3] [4] [5] [6]

中抽取 60 次所得数之和相同。
这盒的平均数是 3.5，SD 是 1.71。和的期望值是 $60\times3.5=210$；和的 SE 是 $\sqrt{60}\times1.71\approx13$。总点数将在 210 附近，相差误差 13。事实上，表 2 中数的和是 212。它偏离期望值大约 $\frac{1}{6}$ 个 SE。

部分(b)是个新问题。取代于把表 2 中 60 个数加起来，我们首先把每一个数归类：它是 6，或者不是 6？（这里只有两类：6 在一类，其它各数都在另一类。）随后我们计算 6 的个数。要注意的一点是，每抛一次，6 的个数要求上升 1，要么保持原来的数。

表 3　计 6 的个数

抛出数	4 5 5 2 4	5 3 2 6 3	5 4 6 2 6	4 4 2 5 6
6 点的连续计数	0 0 0 0 0	0 0 0 1 1	1 1 2 2 3	3 3 3 3 4

表 3 中能看到，它给出了表 2 中前 20 次掷骰子的结果和 6 点出现次数的连续计数。顺计数往前。通常它依然如故，但是每当掷出一个 6 点，它就上升 1。将它更数学化一点：

* 如果掷出 6 点，计数加上 1；
* 如果掷出任何其它各点，计数加上 0。

所以，就机会而言，掷 60 次骰子，出现 6 点的次数恰好与从一个新的盒子

[~~1~~ 0] [~~2~~ 0] [~~3~~ 0] [~~4~~ 0] [~~5~~ 0] [~~6~~ 1]

中抽取 60 次所得数之和相同，理由是：每次掷骰子时，计数在 6 次中有 1 次机会上升 1，6 次中有 5 次机会保持原来的数。同样，在每次抽取时，和的值也是 6 次中有 1 次机会上升 1，6 次中有 5 次机会保持原来的数。计算 6 点的次数就好比从新盒子中抽取数并相加。这使我们有理由使用平方根法则。

新盒子装有 5 个[0]和 1 个[1]。利用捷径方法求得新盒子的 SD 是$\sqrt{1/6\times5/6}\approx0.37$。抽得数之和的 SE 是$\sqrt{60}\times0.37\approx3$。在一颗骰子的 60 次抛掷中，6 点的次数将在 10 附近，加减 3 左右的误差。事实上，在表 2 中有 11 个 6 点：6 点的个数偏离期望值$\frac{1}{3}$个 SE。例题解毕。这是一个新盒子的老故事。

本例得出一个普遍的论点：尽管许多与机会过程有关的问题可能看上去是十分不同的，但是它们能够用同样的方法来解。在这些问题中，随机地从一个盒子中抽取一些票。对抽得数实施一种运算，并且问题要求其结果在给定的区间内的机会。例如：

· 随机放回地从盒子[1] [2] [3] [4] [5] [6]中抽取 60 次。估计抽得数之和在 200 与 225 之间的机会。

这里抽取是作为定量数据——数——来处理的，它们可以相加。对于抽取所实施的运算是加法。

现在有另外一种可能性。例如：

· 随机放回地从盒子[1] [2] [3] [4] [5] [6]中抽取 60 次。估计[6]出现的次数在 10 与 20 之间的机会。

这里抽取是作为定性数据来处理的，(定量数据与定性数据的区分在第 47 页中讨论过。)对抽得数所实施的运算不是加法：它是分类和计数，对每次抽取，我们问：这是 6，还是非 6，随后计算 6 的次数。

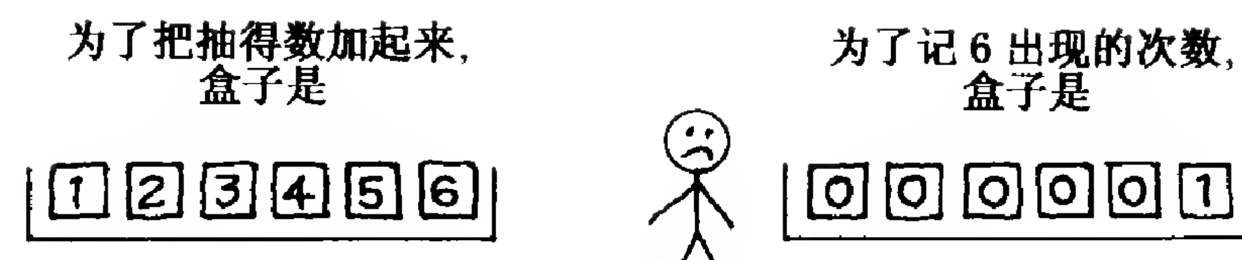
在本章中对抽得数有两种可能的运算：

· 加法

· 分类和计数

但是本节的启示是，两种运算能够以同样的方式来处理——只须改变盒子。

如果你必须对抽得数进行分类和计数，那么在票子上置 0 和 1。那些你要进行计数的票子标上 1，其它票子标上 0。



记住改票子！

例 5. 一枚硬币抛 100 次。求头像数的期望值和标准误差。估计得到的头像数在 40 与 60 元间的机会。

解. 首先是构造一个盒子模型。问题包括把抛得的结果分类成头像还是背面，并随后计头像的次数。因此盒中应当只有 0 和 1。出现头像的机会是 50 对 50，所以盒子应当是 [0 1]。在一枚硬币抛100次中头像数与随机放回地从盒子 [0 1] 中抽取 100 次所得数之和相同。理由是：每次抛硬币都有 50 对 50 的机会使头像数或者增加 1，或者保持不变；同样，每次从盒中抽取都有 50 对 50 的机会使和值或者增加 1，或者保持不变。这完成了模型构造。

由于头像数与抽得数之和相同，平方根法则可以使用，盒子的 SD 为$\frac{1}{2}$。因而 100 次抽得数之和的 SE 为$\sqrt{100}\times\frac{1}{2}=5$。头像数将在 50 附近，加减大约 5 的误差。40 到 60 个头像这个区域表示期望值，加减 2 个 SE 左右误差机会约为 95%，这就完成了解。

要解释这个 95%机会，设想我们在一枚硬币抛 100 次中计算头像出现的次数。你可能得到 44 个头像，再抛 100 次：你可能得到 54 个头像。第三次，头像数将会再度变化，也许为 48 个头像，继续下去，从长远来看，这些计数中大约有 95%出现在 40 到 60 的区域内。

John Kerrich 确实做了这个试验。表 4 给出了这样的结果，把 Kerrich 抛一枚硬币 10 000 次的结果分成 100 个连贯组，每组 100

个。事实上,100 组中 95 组有 40 到 60(包括端点 40 和 60)个头像数。理论看来是有效的。

是到了把平方根法则和平均数律联系起来的时候了。假设抛一枚硬币大数量次。于是头像将在抛的次数的大约一半中出现:

头像数=抛的次数的一半+机会误差

表 4 Kerrich 抛一枚作试验,在抛 100 次的每组中他所得到的头像数。

100 次一组	头像数	100 次一组	头像数	100 次一组	头像数	100 次一组	头像数
1-100	44	2 501-2 600	44	5 001-5 100	42	7 501-7 600	48
101-200	54	2 601-2 700	34	5 101-5 200	68	7 601-7 700	43
201-300	48	2 701-2 800	59	5 201-5 300	45	7 701-7 800	58
301-400	53	2 801-2 900	50	5 301-5 400	37	7 801-7 900	57
401-500	56	2 901-3 000	51	5 401-5 500	47	7 901-8 000	48
501-600	57	3 001-3 100	51	5 501-5 600	52	8 001-8 001	45
601-700	56	3 101-3 200	48	5 601-5 700	51	8 101-8 200	50
701-800	45	3 201-3 300	56	5 701-5 800	49	8 201-8 300	53
801-900	45	3 301-3 400	57	5 801-5 900	48	8 301-8 400	46
901-1 000	44	3 401-3 500	50	5 901-6 000	37	8 401-8 500	56
1 001-1 100	40	3 501-3 600	54	6 001-6 100	47	8 501-8 600	58
1 101-1 200	54	3 601-3 700	47	6 101-6 200	52	8 601-8 700	54
1 201-1 300	53	3 701-3 800	53	6 201-6 300	45	8 701-8 800	49
1 301-1 400	55	3 801-3 900	50	6 301-6 400	48	8 801-8 900	48
1 401-1 500	52	3 901-4 000	53	6 401-6 500	44	8 901-9 000	45
1 501-1 600	54	4 001-4 100	52	6 501-6 600	51	9 001-9 100	55
1 601-1 700	58	4 101-4 200	54	6 601-6 700	55	9 101-9 200	51
1 701-1 800	50	4 201-4 300	55	6 701-6 800	53	9 201-9 300	48
1 801-1 900	53	4 301-4 400	52	6 801-6 900	52	9 301-9 400	56
1 901-2 000	42	4 401-4 500	51	6 901-7 000	60	9 401-9 500	55
2 001-2 100	56	4 501-4 600	53	7 001-7 100	50	9 501-9 600	55
2 101-2 200	53	4 601-4 700	54	7 101-7 200	57	9 601-9 700	50
2 201-2 300	53	4 701-4 800	47	7 201-7 300	49	9 701-9 800	48
2 301-2 400	45	4 801-4 900	42	7 301-7 400	46	9 801-9 900	59
2 401-2 500	52	4 901-5 000	44	7 401-7 500	62	9 900-10 000	52

机会误差可能会有多大?最初,Kerrich 的助手认为它会十分小。记

录表明他是错的。当 Kerrich 继续抛硬币，机会误差按绝对值上升，而相对于抛的次数减少，恰与数学预测相符。（参阅第 296—297 页上图 1 和 2。）

根据平方根法则，机会误差的可能大小是$\sqrt{\text{抛的次数}}\times 1/2$。例如，抛 10 000 次，标准差是$\sqrt{10\ 000}\times 1/2=50$。当抛的次数增加到 1 000 000 时，标准差也增加，不过由于平方根的原因只增加到 500。当抛的次数增加时，头像数的 SE 按绝对值愈来愈大，但相对于抛的次数愈来愈小，这就是头像的比率愈来愈接近于 50%的原因。平方根法则是平均数律的数学解释。

习题 E

1. 一枚硬币抛 16 次。
   (a) 头像数是与随机放回地从下列盒子的一个中抽取 16 次所得数之和相同。是哪一个？为什么？
   (i) | 头像 | 背面 |　(ii) | 0 | 1 |　(iii) | 0 | 1 | 1 |
   (b) 头像数大约是________，加减大约________的误差。
2. 正确还是错误：
   (a) 盒子 | 0 | 0 | 1 | 1 | 1 | 的 SD 是$\sqrt{2\times 3}$。
   (b) 盒子 | −1 | −1 | 1 | 1 | 1 | 的 SD 是$\sqrt{2/5\times 3/5}$。
   (c) 盒子 | 0 | 1 | 1 | 的 SD 是$\sqrt{2/3\times 1/3}$。

   说明之。
3. 随机放回地从盒子 | 1 | 2 | 3 | 4 | 5 | 中抽取 100 次。得到标有数 5 的票子的次数在 8 到 32 之间的机会是多少？
4. 根据最简单的遗传学模型，孩子的性别是随机地确定的，象随机地从盒子

   | 男 | 女 |

   中抽取一张票一样。试问紧接着而出生的 2 500 人（不计双胞胎或其它多胞胎）中女性多于 1 275 人的机会是多少？
5. 在第 330 页的表 4 中应当有多少个计数在 45 到 55 范围内？实际有多少

个？（包括端点）。

6. 一枚硬币抛 10 000 次，头像数将在 4 850 至 5 150 范围内的机会是多少？

7. 一枚硬币抛 1 000 000 次。头像数将在 498 500 至 501 500 范围内的机会是多少？

8. 编制了一个计算机程序去做下面的事。一个装有 10 张空白票的盒子。你告诉程序什么数要写在票上和抽取多少次。然后，计算机将随机放回地从盒中抽取那么多张票，并且把它们的数加起来，打印出和数——但不打印每次抽得数。这程序不知道与抛硬币有关的任何事情。但是，你还是能够用它去模拟抛一枚硬币 1 000 次中的头像数。为什么？

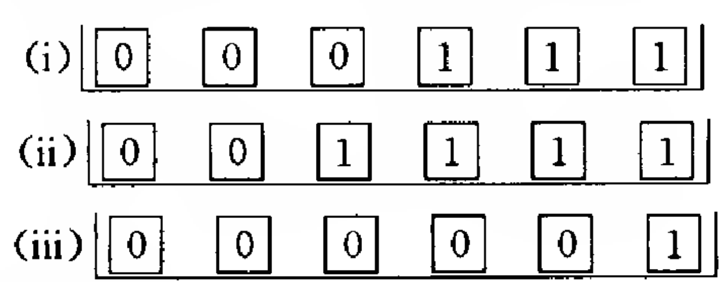
9. 一颗骰子掷 100 次。某人计算幺点的期望数为 $100\times 1/6=16.67$，SE 为 $\sqrt{100}\times\sqrt{1/6\times 5/6}\approx 3.73$。（幺点是 ⊡。）这正确吗？回答是或否，并解释之。

10. 从下面一个盒子中随机地抽取一次。如果你能正确猜出结果，你赢得 1 美元。

(a) 哪一个盒子最有利？哪一个最不利？

(b) 从哪一个盒子抽取变异最小？从哪一个盒子抽取变异最大？

(c) 哪一个盒子有最小的 SD？哪一个有最大的 SD？



这些习题答案在第 705 页上。

6. 复习题

1. 随机放回地从盒子 [1] [6] [7] [9] [9] [10] 中抽取 100 次。

(a) 抽得数之和可能有多小？多大？

(b) 其和在 650 到 750 之间的机会大约为

1%　　10%　　50%　　90%　　99%

说明之。

2. 一赌徒在轮盘赌上赌 100 次，每次都把 1 美元赌注下在列区上。打 1 赔 2，并且 38 次中有 12 次赢的机会。填充下列空白；给出计算过程。

(a)赌 100 次中，赌徒的净利大约为______美元，加减______美元左右的误差。

(b)赌 100 次中，赌徒应该赢______次，加减大约______的误差。

(c)这样下赌注与 Keno 游戏(第 313 页例 1)的单个数字上下赌注作比较会怎样?

3. 匹配数据表与相应的 SD。说明你的理由。

	数据表		SD
(a)	1、−2、−2	(i)	$\sqrt{1/3\times 2/3}$
(b)	15，15，16	(ii)	$2\times\sqrt{1/3\times 2/3}$
(c)	−1，−1，−1，+1	(iii)	$3\times\sqrt{1/3\times 2/3}$
(d)	0，0，0，1	(iv)	$\sqrt{1/4\times 3/4}$
(e)	0，0，2	(v)	$2\times\sqrt{1/4\times 3/4}$

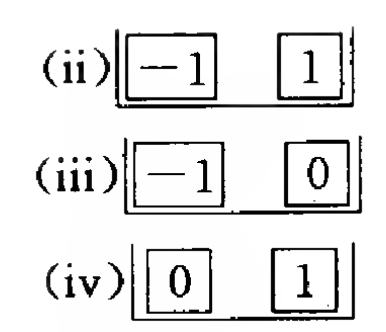
4. 一大群人聚在一起，每人掷一颗骰子 180 次，并且计数 [•] 出现的次数。这些人中大约有百分之几所得的计数应当在 15 到 45 的范围内?

5. 一颗骰子掷一定次数，目的是猜测它们的总点数。对猜测偏离的点数，每点要处以 1 美元的罚款。例如，如果你猜 200，而实际掷出的总点数为 215，你就输掉 15 美元：你宁愿要哪一种：掷 50 次，还是 100 次？解释之。

6. 一枚硬币抛 100 次。

(a)“头像数一背面数”的差与从下面盒子之一中抽取 100 次所得数之和相同。是哪一个盒子，为什么?

(i) | 头像 背面 |



(b)求此差的期望值和标准误差。

7. 一枚硬币抛 1 000 次。

(a)如果有 528 个头像，硬币出现头像次数的百分数是多少?

(b)如果硬币出现头像数为 43.7%，有多少个头像?

(c)有两种选择：

(i)如果头像数在 490 和 510 之间，赢 1 美元。

(ii)如果头像数的百分率在 49%和 51%之间，赢 1 美元。

哪一种选择较有利？或它们是相同的吗？说明之。

8. 随机放回地从盒子 [1] [2] [3] [4] [5] [6] 中抽取 100 次。

(a)如果抽得数之和是 310，其平均数是多少?

(b)如果抽得数的平均数是 4.3，其和是多少?

(c)估计抽取的平均数在 3 与 4 之间的机会。

9. 一赌徒在轮盘赌上赌 1 000 次，有两种可能性：

(i)每次把 1 美元赌注下在列区上。

(ii)每次把 1 美元赌注下在 1 个数上。

下赌注在列区上打 1 赔 2，38 次中有 12 次赢的机会；下赌注在一个数上打 1 赔 35，38 次中有 1 次赢的机会。正确还是错误：

(a)对(i)与(ii)来说最终赢钱的机会是一样的。

(b)赢 100 美元以上的机会，(ii)较大。

(c)输掉 100 美元以上的机会，(ii)较大。

说明之。

10. 某多项选择测验有 25 道题，每一道题有 5 种可能的答案，其中一个是正确的。每一个正确答案给 4 分，每一个错误答案扣 1

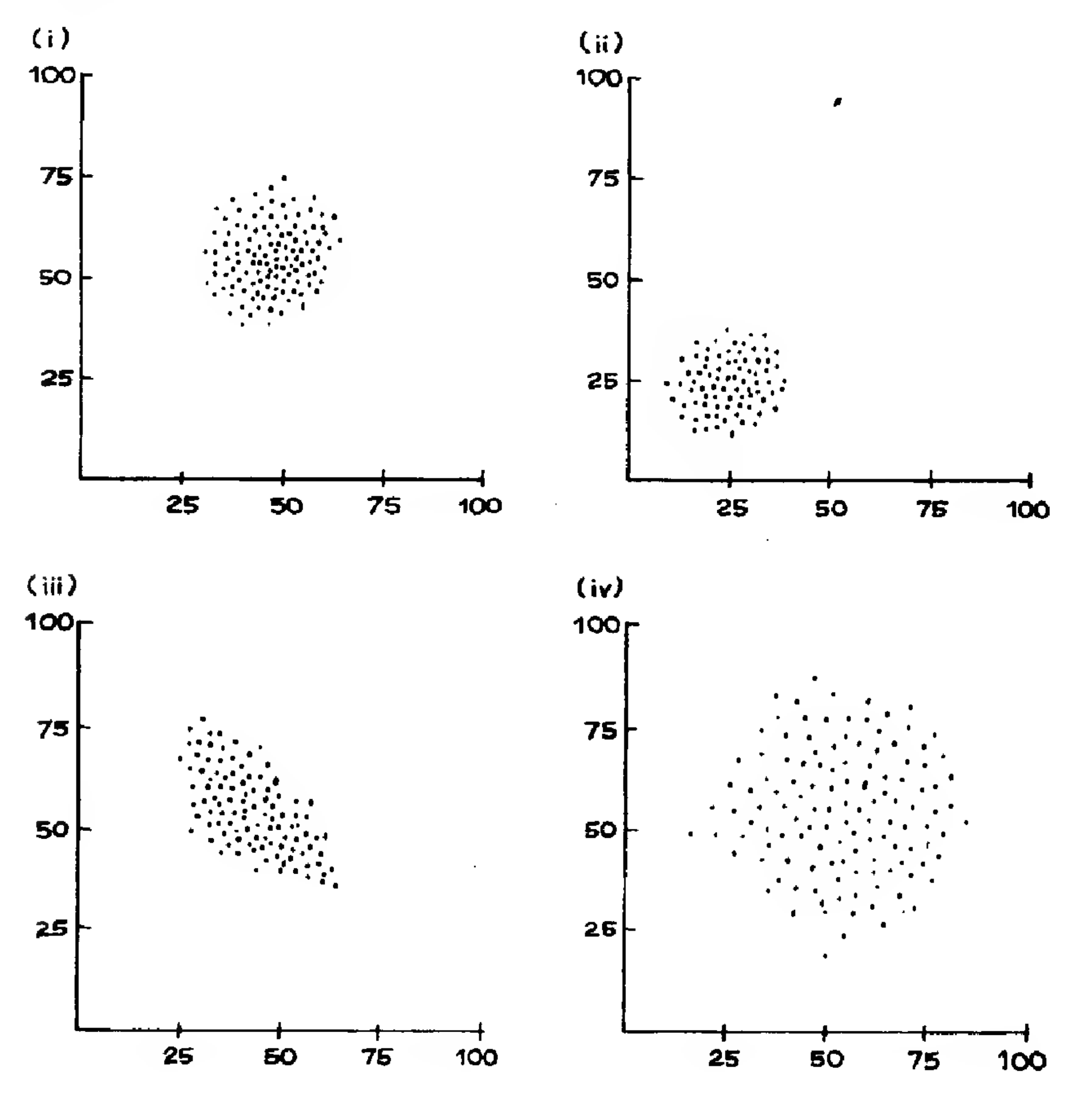
分。通过测验的分数是 30。

(a)如果一学生随机地回答所有问题，通过测验的机会是多少？

(b)只经过猜测，学生能够获得 10 分或者 10 分以上吗？

回答是或否，并说明理由。

11. 设想画一张第 330 页上表 4 的散点图如下。画一点，其 x 坐标是抛编号为 #1—100 中的头像数，y 坐标是抛编号为 #101—200 中的头像数。这给出一点(44,54)。随后画一点，其 x 坐标是编号为 #201—300 中的头像数，y 坐标是编号为 #301—400 中的头像数。这给出一点(48,53)，如此等等。散点图看上去像(i)、(ii)、(iii)还是(iv)？说明之。



7. 附言

本章的习题给人们上了一堂令人不快的课：你赌得愈多，你就输得愈多。基本原因是：在你的期望净利都是负的意义下，一切打赌都是不公平的。所以平均数律只对庄家起作用，而不是对你。当然，本章只讨论了一些简单的对策，赌徒们在轮盘赌、掷两颗骰子的赌博以及类似的其它赌博中发展了下赌注的复杂系统。但是没有一种调整不公正的打赌系统能使得你的期望净利变成正的，这是一个数学定理。要证明这一定理，只需要两个假设条件：

- 你不是一个具有超人洞察力的人；
- 你的财政来源是有限的。

二十一点的赌博是一个例外，在这种赌博中某些情况下存在一些具有正期望净利的打赌[3]。其结果是，人们在二十一点赌博中赢得过许多钱。

8. 小结

1. 由机会过程产生的数应当在期望值附近某个地方，但偏离一个机会误差。这个机会误差的可能大小由标准误差给出。例如，从盒子中抽得数之和将在期望值的附近，加减大约一个标准误差。

2. 当随机放回地从装有标上数字的票子的盒中抽取时，每次抽取都给和增加一个大约等于盒平均数的量。所以和的期望值是

$$(\text{抽取次数})\times(\text{盒平均数})。$$

3. 当随机放回地从装有标上数字的票子的盒中抽取时，

$$\text{和的 SE}=\sqrt{\text{抽取次数}}\times(\text{盒子 SD})。$$

这是平方根法则。

4. 当盒中票只标有两个不同数字时，盒子的 SD 可用捷径方法计算：

$$(\text{大的数}-\text{小的数})\times\sqrt{\frac{\text{具有大数的}}{\text{票数占的比例}}\times\frac{\text{具有小数的}}{\text{票数占的比例}}}$$

5. 如果你需要把抽得的票分类和计数，记住在你要计数的票上置 1 和其它票上置 0。

6. 假如抽取的次数充分大，正态曲线能用来计算抽得数之和的机会。

18
概率直方图的正态近似

每个人都相信[正态近似],试验者出于,他们想这是一个数学定理,而数学家出于,他们想这是一个实验事实。

——G. Lippmann(法国物理学家,1845—1921)

1. 引言

按照平均数律,当一枚硬币被抛大数量次时,头像的比率将接近于50%。1700年前后,瑞士数学家James Bernoulli将这一点置于严格的数学基础之上。20年以后,Abraham de Moivre通过指出如何计算头像百分率落入围绕50%的一个任意给定区间的机会,在贝努利的工作上做了实质性的改进。计算不是精确的,但是随着抛的次数增加,近似会愈来愈好。(De Moivre的计算在前面第13章中讨论过)。

Bernoulli和de Moivre两人关于硬币都作了同样的假设:每次抛都是独立的,而且每次抛时硬币出现头像与出现背面是等可能的。从这些假设出发可以得出,钱币以正反面的任何一种特定型式呈现与以任何其它的型式呈现,其可能性是一样的。Bernoulli所做

的是揭示了，对大多数型式，其中约有 50%的记录是头像。

你能看到，这种情况甚至在抛 5 次中就开始出现。设想一枚硬币抛 5 次，并记录每次它出现的状态。一种可能型式是全部都出现头像：

H H H H H

这一型式有 5 个头像。有 4 个头像的型式有多少呢？答案是 5 种：

THHHH HTHHH HHTHH HHHTH HHHHT

例如，型式 THHHH 意味着在第一次抛时出现背面，随后接连给出 4 个头像，表 1 指出任何一个给定头像数有多少种型式。对抛 5 次而言，硬币能呈现的可能型式总共有 $2^5=32$ 种，并且在 32 种型式中有 20 种的头像数接近一半（5 中有 2 或者 3 个头像）。

表 1 一枚硬币抛 5 次中对应于已知头像数的型式数

头像数	型式数
0	1
1	5
2	10
3	10
4	5
5	1

对任何抛硬币的次数，de Moivre 在一个小的误差界限内，设法算出具有给定头像次数的型式数。为了弄清其含义，以抛币 100 次为例。他必须考虑的型式数大约是 2^{100}。这是一个十分巨大的数。如果你试图写出全部这些型式，在本书大小的一页上大概只能写 100 种型式，在你写完这些型式的时侯，你的书足以塞满一个从地球到最远的已知星球的大书架。

尽管如此，数学家们有一个计算正好有 50 个头像的型式数的公式，由二项系数（第 15 章，除非你跳过这一内容，在这里不会有什么困难）给出：

$$\frac{100!}{50!\ \times 50!}=\frac{100\times 99\times \cdots\cdots \times 51}{50\times 49\times \cdots\cdots \times 1}$$

但是，这公式对 de Moivre 没有直接帮助，因为它没有说出这个数有多大。计算看上去是令人十分不愉快的，而用计算器[①]可算出这个数为 1.01×10^{29}。此外 2^{100} 是 1.27×10^{30}，所以在一枚硬币抛 100 次中恰好得到 50 个头像的机会是

$$\frac{1.01\times10^{29}}{1.27\times10^{30}}\approx0.08=8\%$$

当然，de Moivre 没有任何象现代计算机那种可使用的东西。他需要一种不用算术计算的数学方法用以估计二项系数。他找到了实施的方法(尽管近似公式通常归功于另一位数学家 James Stirling)。这种估计导致 de Moivre 发现正态曲线。例如，他发现一枚硬币抛 100 次中恰好得到 50 个头像的机会大约等于正态曲线下 -0.1 到 $+0.1$ 之间的面积，事实上，他做到了从数学上证明，头像数的整个概率直方图，在抛的次数增加时愈来愈接近于正态曲线。当代研究工作者们已经把这一结果推广到随机地从任何一个装有票的盒子中抽取所得数之和上。

De Moivre 的论证细节在这里引入是太复杂了，但是可以用计算机作图来图示地追溯他的步骤。

2. 概率直方图

当一个机会过程产生一个数时，期望值与标准误差是该数将在什么地方的一个指南。但是概率直方图给出一张完整的图

> 概率直方图是新的一类图。它表示机会而不是数据

下面用一个具体例子说明这一思想。赌徒玩一种双骰子赌博，把赌注下在一对骰子显示的点数之和上。所以输赢依赖于每次掷出的可能和 2 到 12，为了得到这些机会，庄家可以雇一个人去掷一对骰子，并且求出总点数。这一试验在计算机上重复模拟 10 000 次，前 100 次重复的结果列出在表 2 中。

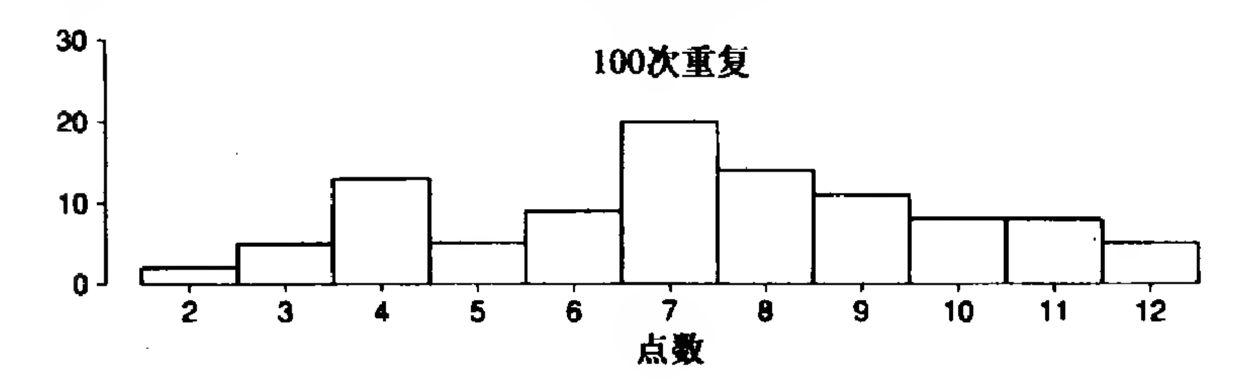
表 2　掷一对骰子。计算机模拟掷一对骰子并求出它们点数和。重复这一过程 10 000 次，表出前 100 次重复结果。

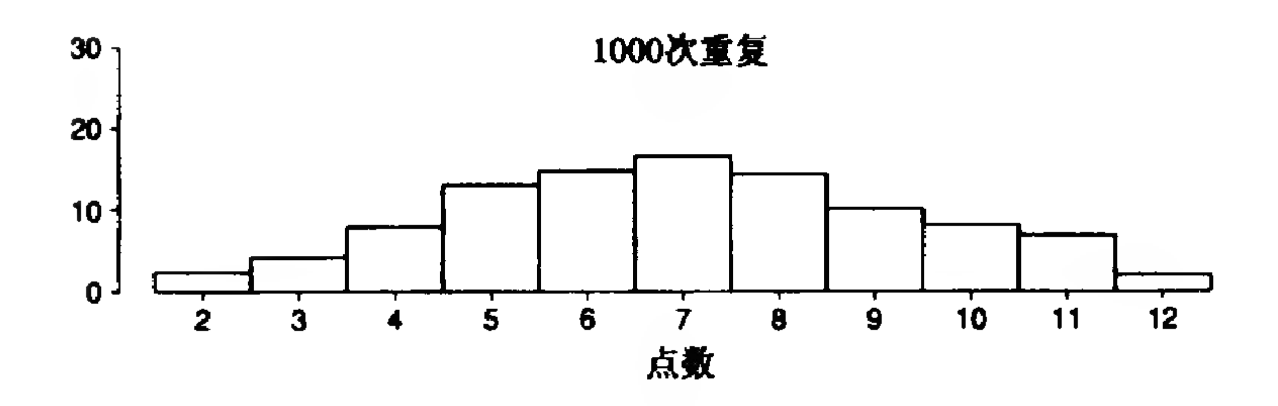
重复号	和	重复号	和	重复号	和	重复号	和	重复号	和
1	8	21	10	41	8	61	8	81	11
2	9	22	4	42	10	62	5	82	9
3	7	23	8	43	6	63	3	83	7
4	10	24	7	44	3	64	11	84	4
5	9	25	7	45	4	65	9	85	7
6	5	26	3	46	8	66	4	86	4
7	5	27	8	47	4	67	12	87	7
8	4	28	8	48	4	68	7	88	6
9	4	29	12	49	5	69	10	89	7
10	4	30	2	50	4	70	4	90	11
11	10	31	11	51	11	71	7	91	6
12	8	32	12	52	8	72	4	92	11
13	3	33	12	53	10	73	7	93	8
14	11	34	7	54	9	74	9	94	7
15	7	35	7	55	10	75	9	95	7
16	8	36	6	56	12	76	11	96	9
17	9	37	6	57	7	77	6	97	10
18	8	38	2	58	6	78	9	98	5
19	6	39	6	59	7	79	9	99	7
20	8	40	3	60	7	80	7	100	7

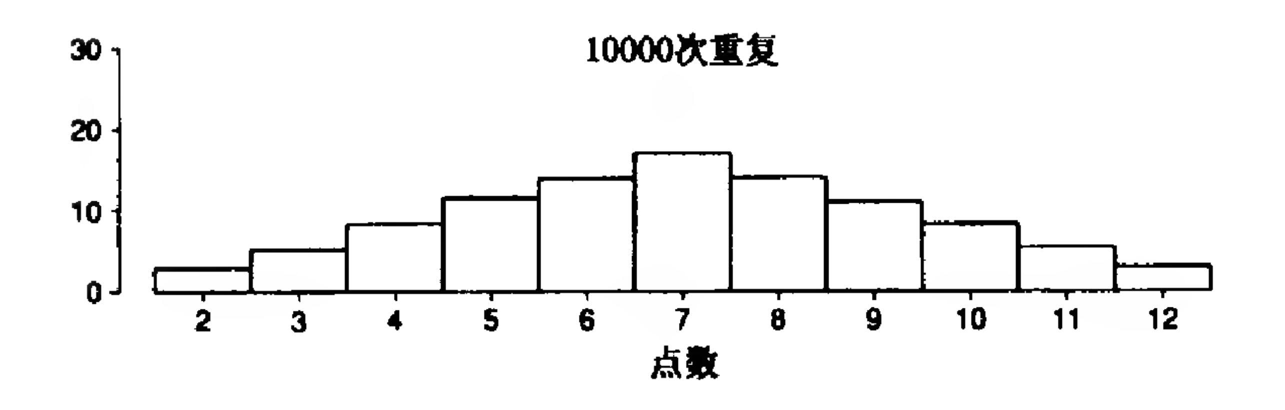
图 1 中第一部分表示表 2 中数据的直方图。和数为 7 出现 20 次，所以在 7 上的矩形有面积 20%，而且类似地可对其它可能的和作出，第二部分表示前 1 000 次重复的经验直方图，第三部分是全部 10 000 次重复的经验直方图。这些经验直方图收敛于图的最后一部分所表示的理想概率直方图。(经验的意思是"实验观察到的")

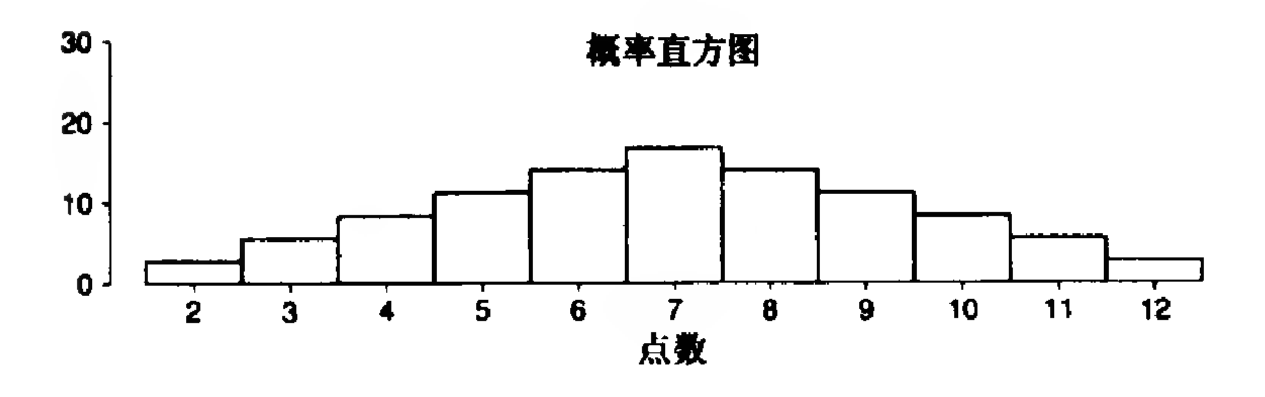
当然，图 1 中的概率直方图能够通过理论论证求得。如在第 14 章中所给出的，在 36 次中有 6 次机会掷出和为 7。这是 $16\frac{2}{3}\%$ 机会，因此，概率直方图中 7 上的矩形面积等于 $16\frac{2}{3}\%$。可以对其

图 1　经验直方图收敛于概率直方图。计算机模拟掷一对骰子并求出点数和,重复过程 100 次并作出 100 个和的直方图(顶端部分)这是一个基于观察值的经验直方图,第二部分是重复 1 000 次的,第三部分是重复 10 000 次的,(每次重复是掷一对骰子)最下面一部分是掷一对骰子时点数和的最终设想或者概率直方图。









它矩形进行类似讨论。

> 概率直方图用面积表示机会。直方图是由矩形组成的，对于抽得数之和，每一个矩形的底是以可能值为中心；矩形的面积等于获得该值的机会[3]。直方图的总面积是 100%

举另一个例子，考察一对骰子的数字之乘积而不是数字之和，编制计算机程序去一次又一次地重复如下的机会过程：掷一对骰子，并求它们所显示的两个数之积。图 2 的顶端部分给出 100 次重复的经验直方图。乘积 10 出现 4 次，所以 10 上方的矩形面积等于 4%；对其它乘积值用相同方法处理。第二部分给出 1 000 次重复的经验直方图；第三部分是关于 10 000 次重复的。(每一次重复包括掷一对骰子以及取它们的点数之积。)最后部分表示概率直方图。10 000 次重复的经验直方图看上去几乎完全与概率直方图一样。

图 2 十分不同于图 1：新的直方图有间隙。考虑乘积的可能值有助于了解其中的原因。如果两颗骰子都出现⚀，最小乘积值是 1；如果两颗骰子都出现⚅，最大乘积值是 36，但是没有办法得到乘积为 7。因此在 7 上方没有矩形；机会是 0。根据同样的理由，在 11 上方也没有矩形。所有间隙都能这样解释。

做练习将会帮助你学会如何看懂概率直方图，这些直方图将用于本书的余下的部分。

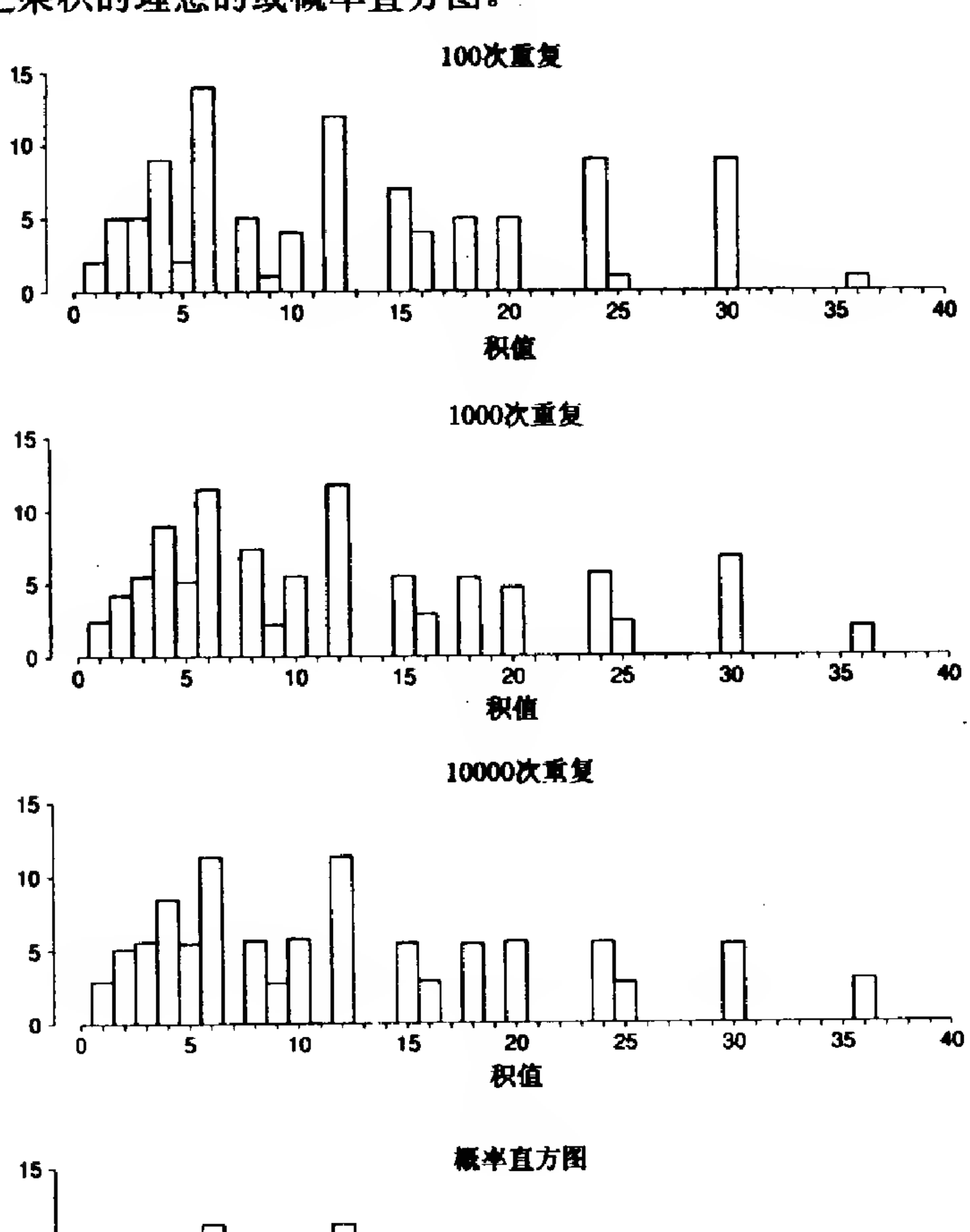
习题 A

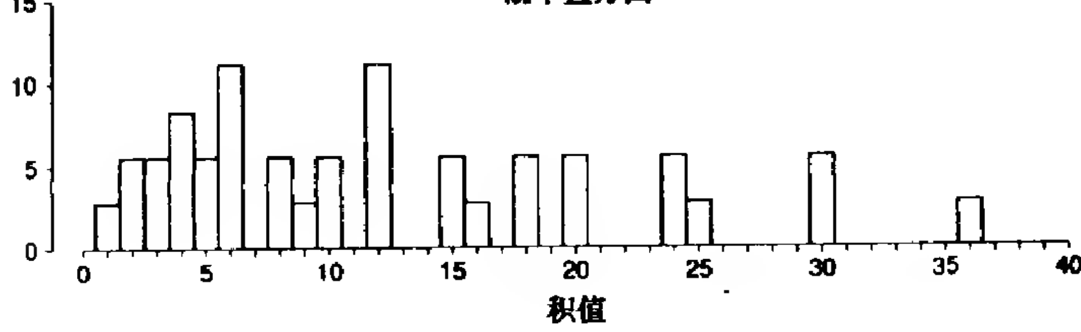
1. 下面的图是一张从盒子 [1] [2] [3] [4] [5] 中抽取 25 次所得数之和的概率直方图，阴影区域的面积表示和在____和____之间(包括端点)的机会(见 345 页上图)。
2. 本题参照第 342 页上图 1
   (a) 如果掷一对骰子，点数和最可能是____。
   (b) 一对骰子掷 1 000 次，最经常出现的和是什么?
   (c) 在图 1 的顶上部分，4 上方的矩形较大于 5 上方的矩形。这是由于 4 比

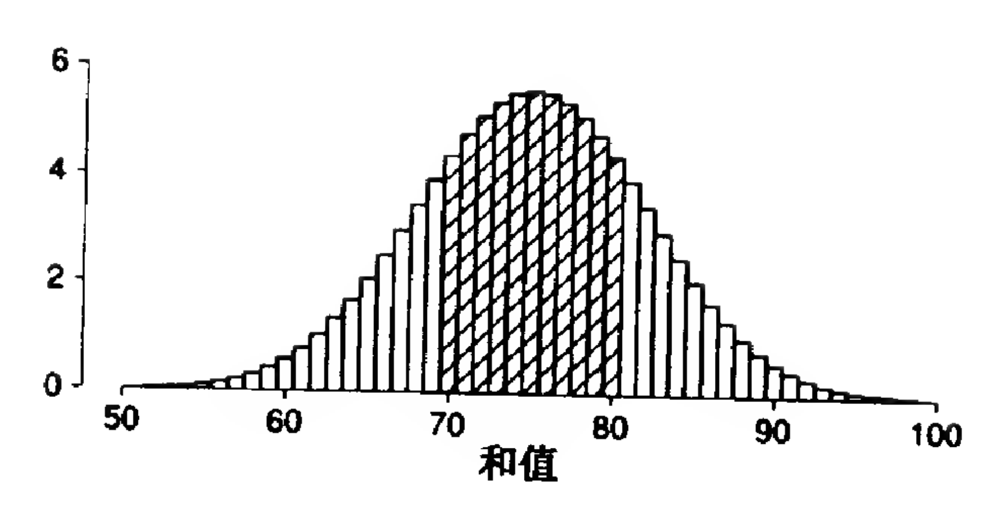
5 更可能出现吗？说明之。

(d)看一下此图的顶上部分。8 上方的矩形面积表示

图 2 经验直方图收敛于概率直方图。计算机模拟掷一对骰子和取它们所显示两数之乘积。重复这过程 100 次，并作 100 个乘积的直方图(顶端部分)。这是一个基于观察值的经验直方图，第二部分是重复 1 000 次的经验直方图，第三部分是重复 10 000 次的。(每次重复是掷一对骰子。)最后部分是掷一对骰子所显示的两数之乘积的理想的或概率直方图。







(i)掷一对骰子得到总数为 8 点的机会。

(ii)一对骰子掷 100 次得到总数为 8 点的机会。

(iii)表 2 中出现总数为 8 的次数的百分数。

选择一种,并说明之。

3. 第 344 页图 2 是关于一对骰子出现的数之乘积的。

(a)如果骰子出现　　　　,积是多少? 如果它们出现　　　　,积是多少?

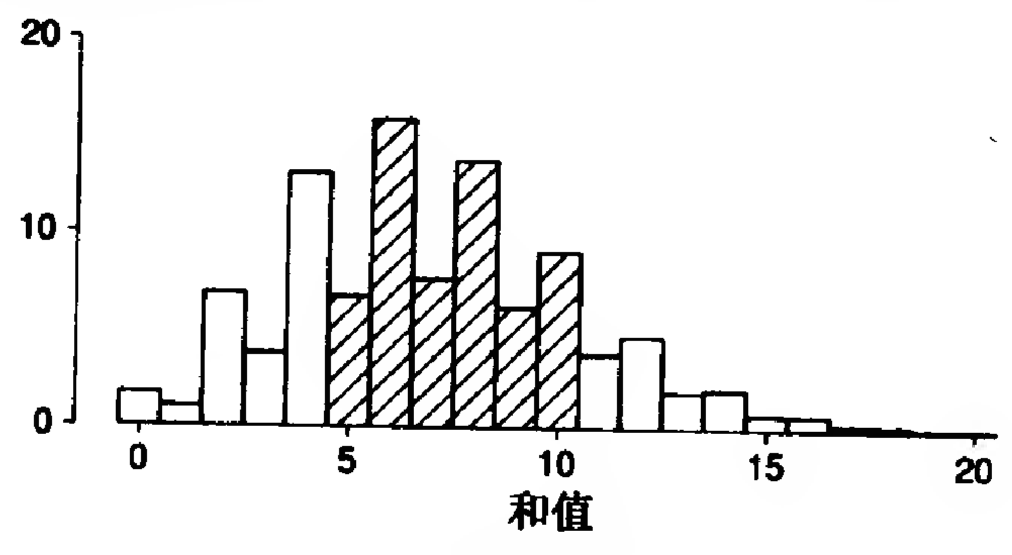
(b)"2 与 3 作为乘积值是等可能的",你应该看图中哪一部分以检验这一陈述? 它正确吗?

(c)一对骰子 1 000 次投掷中,哪个乘积值出现更经常些:是 2 还是 3? 说明之。

(d)无一直方图在 14 上方有矩形,为什么?

4. 看图 2 中最下面部分,在 6 上方的矩形面积是 11.1%,这 11.1% 表示什么?

5. 下图是一张随机放回地从一个盒子中抽取 25 次所得数之和的直方图。



正确还是错误:阴影面积表示你抽得的数在 5 与 10 之间(包括端点)的次数的百分率。

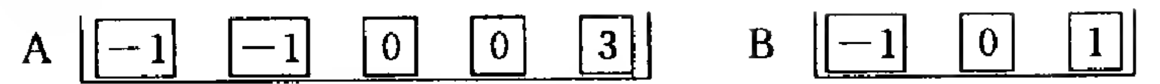
6. 下图表示随机放回地分别从盒子 A 和 B 中抽取 25 次所得数之和的概率

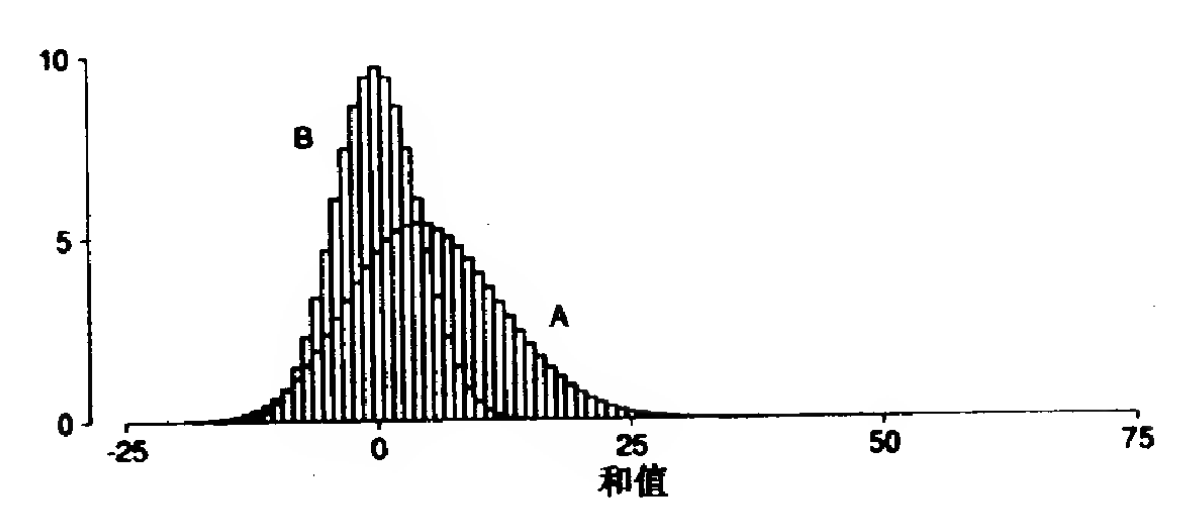
直方图。

(a)哪一个和倾向于为较大值?

(b)哪一个和较容易预测?

说明之





这些习题答案在第 705—706 页上。

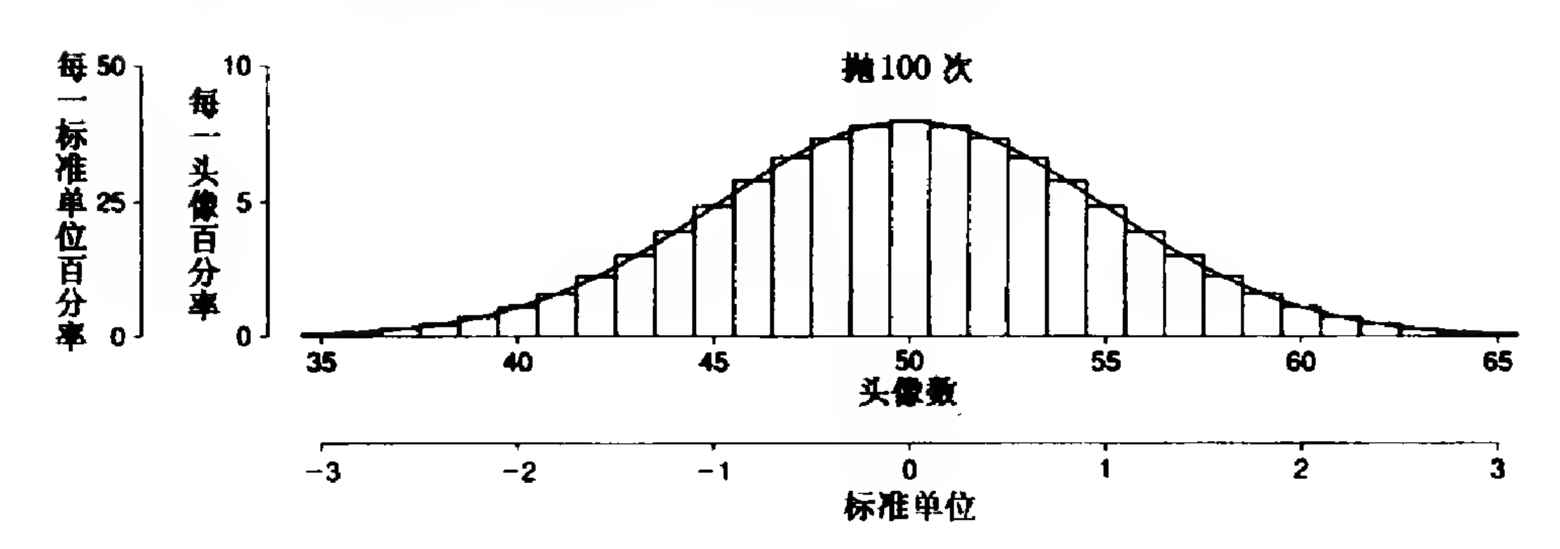
3. 概率直方图与正态曲线

本节的目的是讲述在抛一枚硬币的次数很大时,头像数的概率直方图接近于正态曲线,例如,假定一枚硬币抛 100 次,头像数的概率直方图更为锯齿状,但它仿效正态曲线十分好:参阅图 3。

此图有两根水平轴。概率直方图是相对于上面一根轴绘制的,该轴标明的是头像数。正态曲线是相对于下面一根轴描绘的,该轴标明的是标准单位。头像数的期望值是 50,且 SE 是 5。因此头像数轴上的 50 对应于标准单位轴上的 0;55 对应于+1,等等。

图中也有两根垂直轴。概率直方图是相对于里边一根表示每个头像数的百分率的轴绘制的。正态曲线是相对于外边一根表示每标准单位的百分率的轴描绘的。为知道尺度如何匹配,取顶端值,SE 是 5,所以有 5 个头像对 1 个标准单位。于是每一标准单位的百分率 50 对应每一头像数的百分率 10:50/5=10。用同样的

图 3　在一枚硬币抛 100 次中头像数的概率直方图与正态曲线的比较。曲线是按直方图的标准单位尺度绘制的。



方法能够处理任何其它各对的值。

图 4 分别给出一枚硬币抛 100、400 和 900 次中头像数的概率直方图。抛 100 次，直方图遵循正态曲线，不过有较多的锯齿形。抛 900 次，直方图实际上已经与曲线没有区别。在十八世纪初，de Moivre 通过纯粹的数学推理证明了收敛性必定发生。

4. 正态近似

在第 17 章中已经用正态曲线来计算机会。本节将阐述它的逻辑原理。还将提出在抛的次数较少或者要求额外精确度时，应当使用的一种处理端点的技巧。

例1. 一枚硬币抛 100 次。估计得到头像数

(a)在 45 与 55 之间(包括端点)

(b)在 45 与 55 之间(不包括端点)

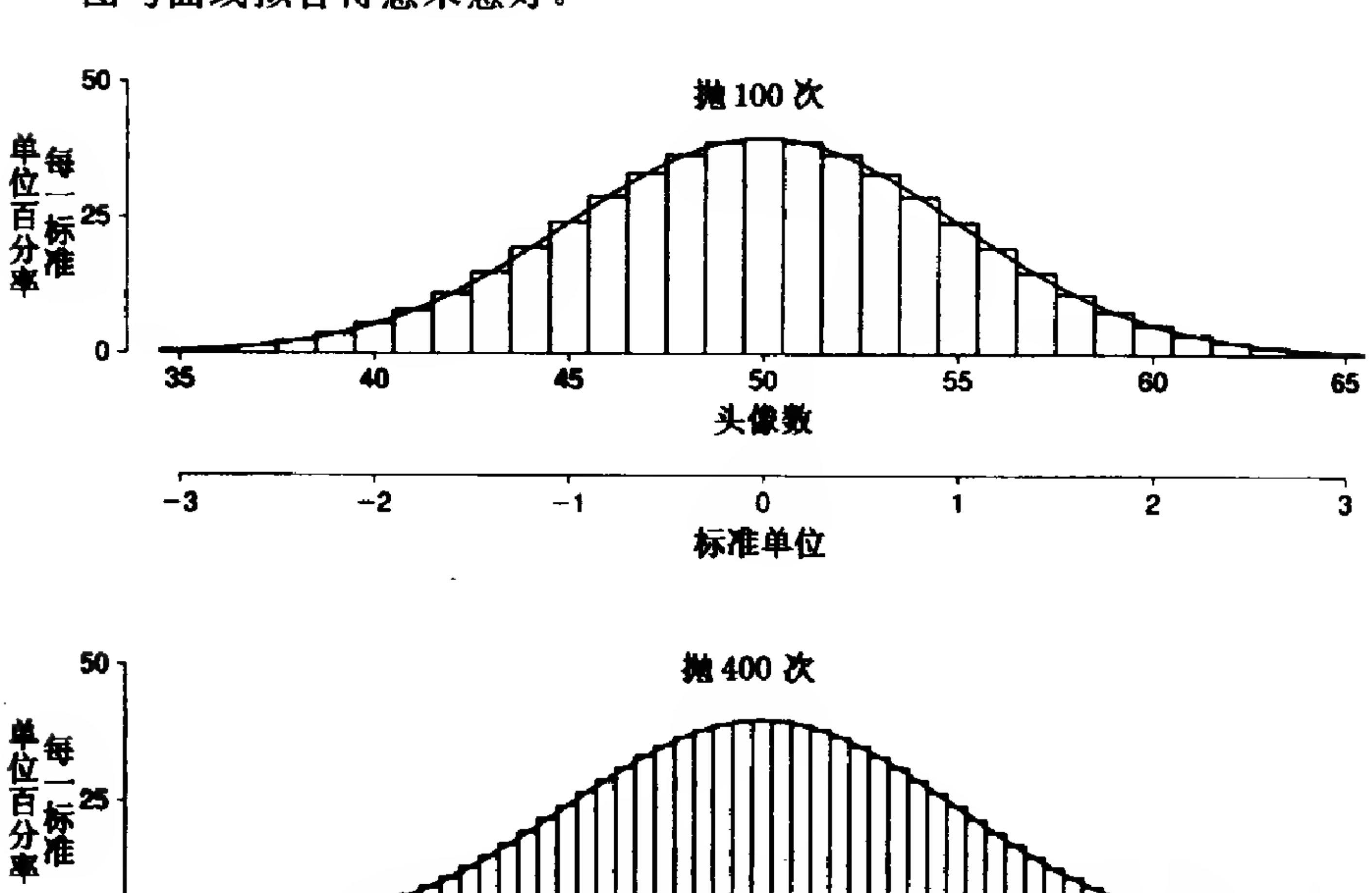
(c)恰好为 50

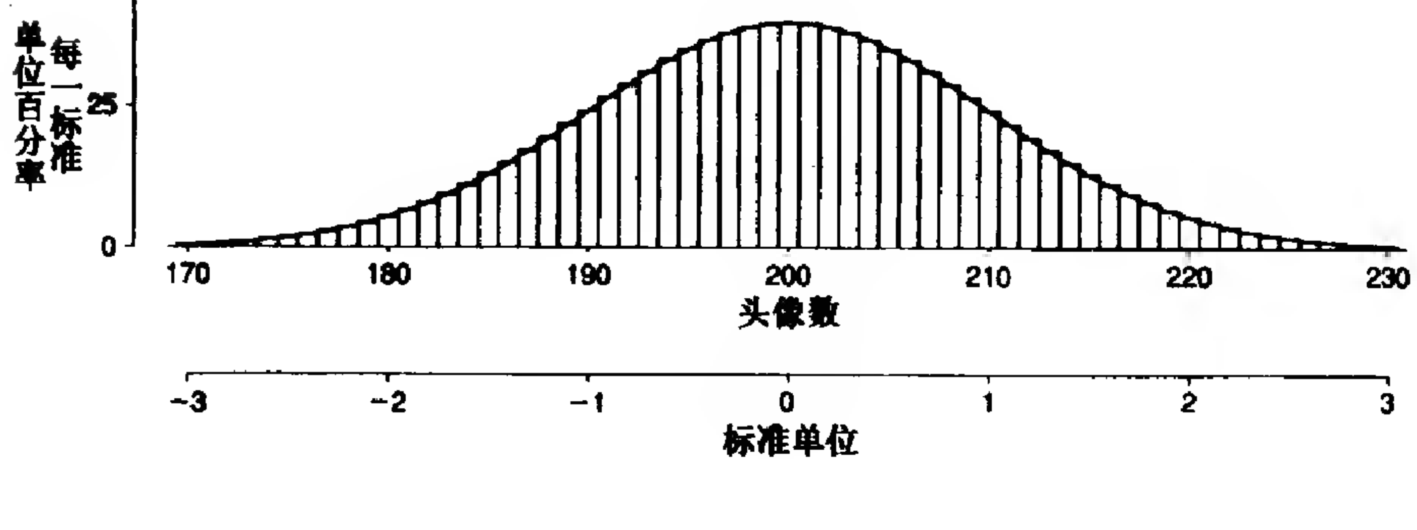
的机会

解：如在第 329 页所指出的，头像的期望值是 50，标准误差是 5。

(a)得到 45—55 个头像(包括端点)的机会等于图 3 中值 45—55 上方 11 个矩形的面积总和。这些矩形复制在图 5 中。

图 4　正态近似，给出一枚硬币抛 100、400 和 900 次中头像数的概率直方图。为了比较也给出正态曲线。随着抛的次数增加，直方图与曲线拟合得愈来愈好。





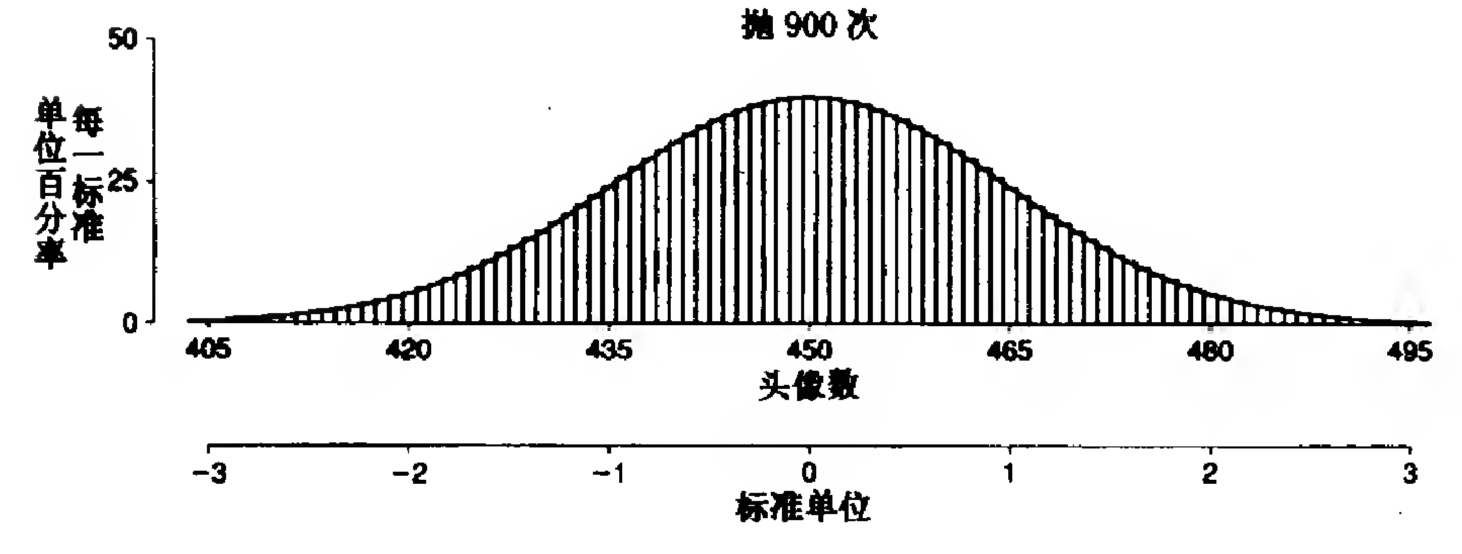
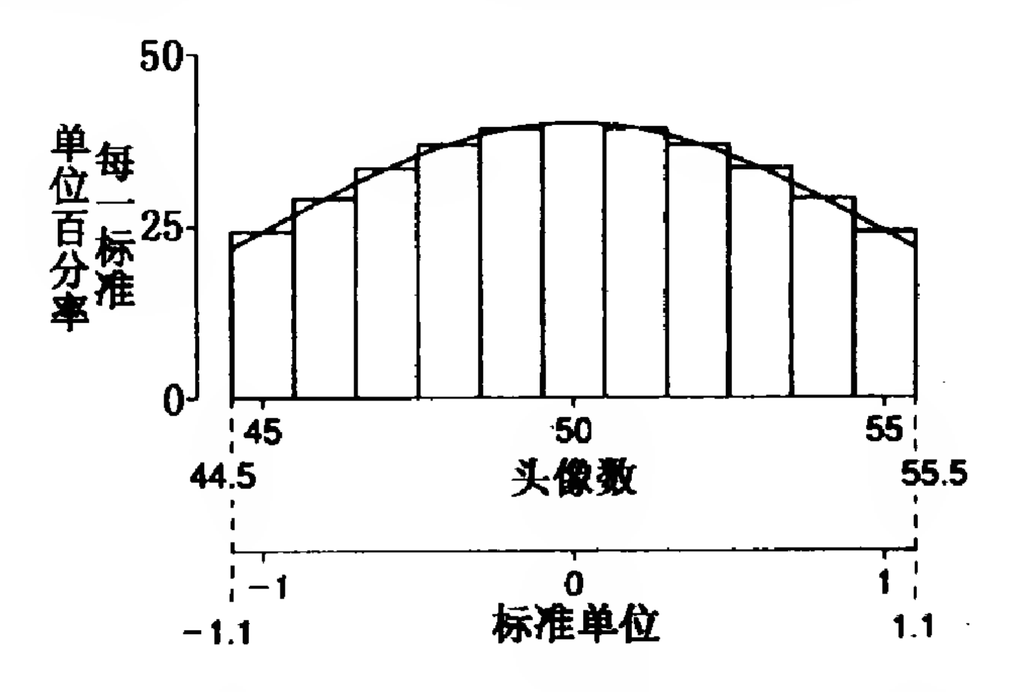
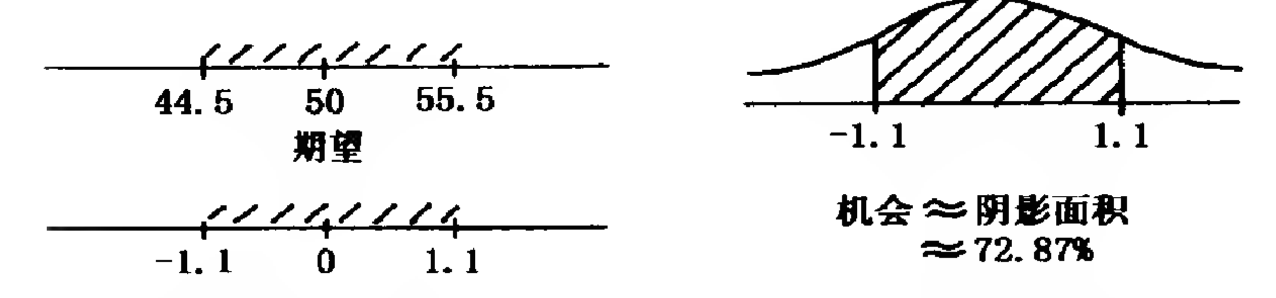


图 5 一枚硬币抛 100 次中头像数的概率直方图。在抛 100 次中得到 45—55 个头像(包括端点)的机会恰好等于在头像数尺度下 44.5 与 55.5 之间,相应于标准单位尺度下 −1.1 与 1.1 之间,直方图下的面积。由于直方图与正态曲线拟合得如此紧密,这机会近似地等于在 −1.1 与 1.1 之间正态曲线下的面积。

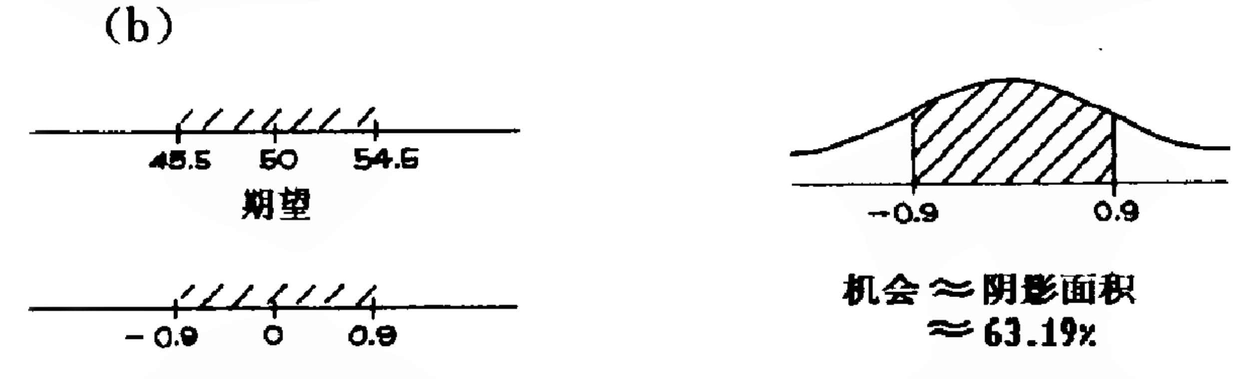


我们欲求在头像数尺度下 44.5 和 55.5 之间直方图下的面积。由于此直方图拟合正态曲线如此紧密,这就几乎等于曲线下的面积。



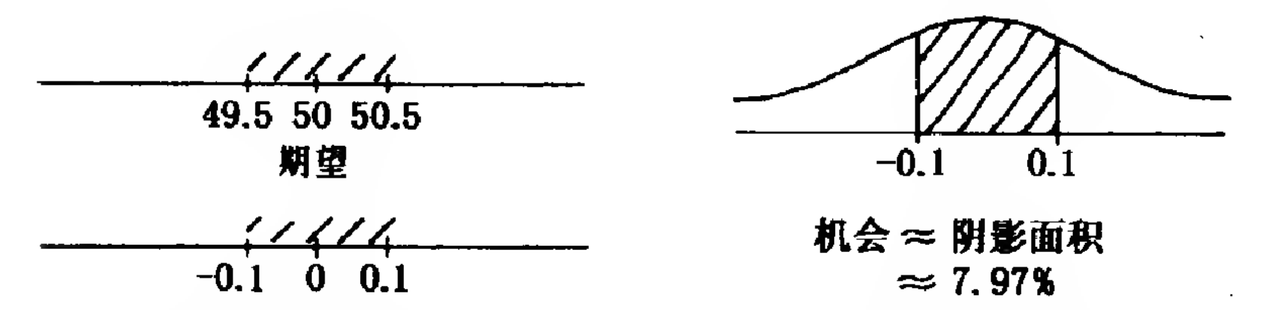
实际机会精确到两位小数,是 72.87%[4],近似值丝毫不差。

(b)



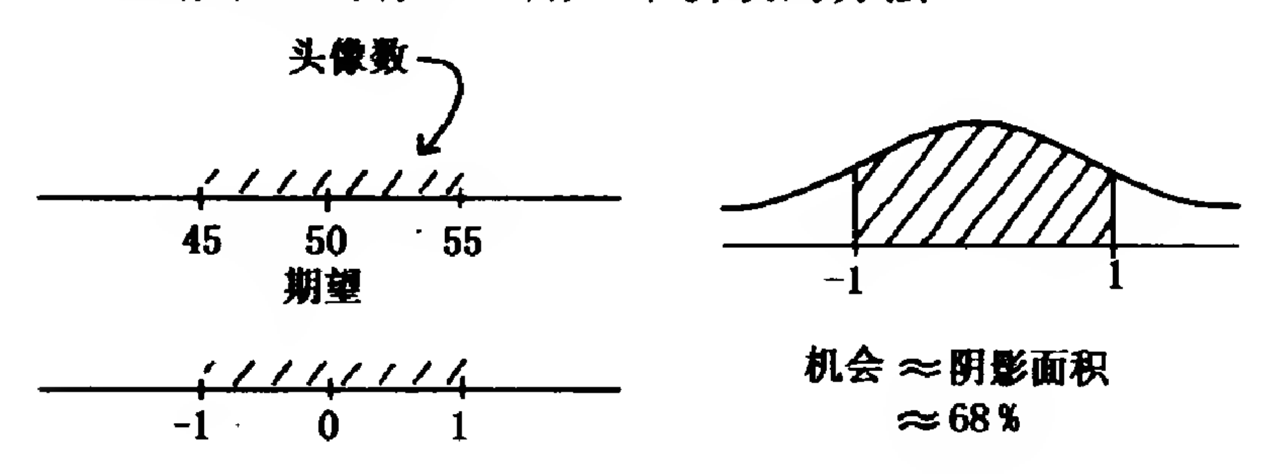
得到头像数在 45 到 55 之间(不包括端点)的机会等于在值 46 到 54 上方 9 个矩形的面积总和。那是在头像数尺度下 45.5 到

54.5 之间，相应于标准单位尺度下－0.9 到 0.9 之间直方图下的面积。这机会几乎等于－0.9 与 0.9 之间正态曲线下的面积 63.19%（实际机会精确到二位小数，是 63.18%。）



恰好得到 50 个头像的机会等于 50 上矩形的面积，这矩形的底，按头像数尺度，是从 49.5 到 50.5，按标准单位尺度那就是－0.1到 0.1。因此，就机会近似地等于正态曲线下－0.1 到 0.1 之间的面积。从正态分布表查得面积大约是 7.97%（而实际机会精确到两位小数是 7.96%）例题解毕。

通常，问题只是求（譬如说）头像数在 45 与 55 之间的机会，而没有规定端点，此时你可以用一个折衷的方法：



这等同于将 45 与 55 之间直方图下的面积用这两个值以标准单位之间正态曲线下的面积所代替。它把两个端点的矩形分成两半，而没有给出如例中所用方法一样高的精度。如果矩形很大，或者要求的精度很高，留意端点的归属是值得的。

正态近似是在计算面积之前用正态曲线去替换实际概率直方图，这种替换只有在概率直方图遵循正态曲线时才是合理的，近似的意义是，概率直方图常常难于计算出，而正态曲线下的面积却很容易在表中查到。

习题 B

1. 图 5 中，得到 52 个头像的机会恰好等于在______与______之间____下的面积。填充空白。对于最后一个空白，你可以选择：正态曲线、概率直方图。说明你的答案的理由。
2. 一枚硬币抛 100 次，估计得到 60 次头像的机会。
3. 从 Kerrich 抛 10 000 次硬币的数据中可以紧挨着取 100 个为一组（第 330 页表 4）。大概应当有多少组正好出现 60 次头像？实际有多少组的头像数是 60？
4. 一枚硬币抛 10 000 次，估计得到头像数在

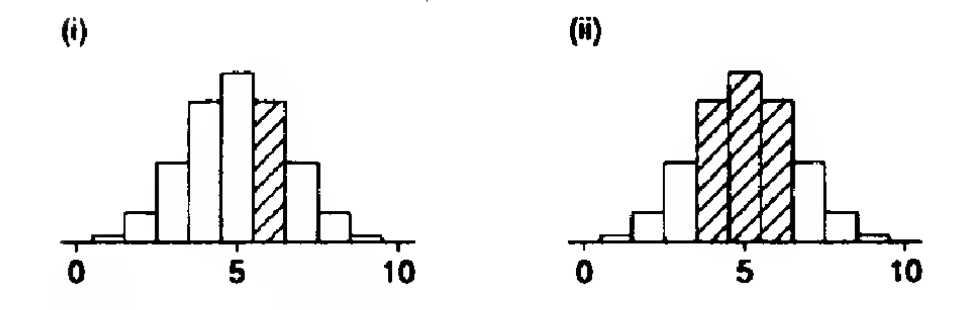
   (a)4 900 到 5 050 之间

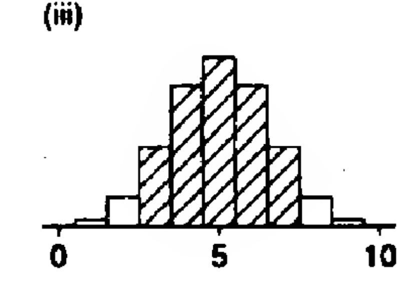
   (b)4 900 或少于 4 900

   (c)5 050 或多于 5 050

   的机会
5. (a)假如你打算估计在一枚硬币抛 100 次中得到或少于 50 次头像的机会。你应当留意矩形的边界吗？

   (b)对于抛 900 次中得到或少于 450 次头像，进行与(a)中一样的讨论，不需计算，只要观察图 4。
6. 一枚硬币抛 10 次，下面用三个不同的阴影面积表示头像数的概率直方图，一个对应于得到头像 3 到 7 次（包括端点）的机会。一个对应于得到头像 3 到 7 次（不包括端点）的机会。再一个对应于正好得到 6 次头像的机会。哪一张图表示哪一种机会，为什么？





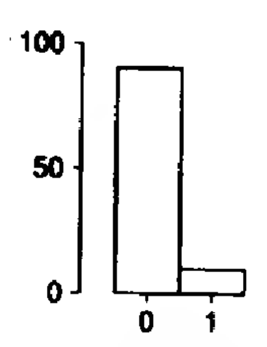
这些习题的答案在第 706—707 页上。

5. 正态近似的范围

到目前为止，讨论只与一枚有 50%机会出现头像或背面的硬币有关。有关从盒中抽取的情形该是怎样的呢？你只要记住一件

事，正态近似就完全一样起作用。盒中诸数的直方图偏离正态曲线越大，在取近似之前抽取数就需要越多。例如，盒子|9 张[0] [1]|。正如票的直方图(图 6)所示那样，这个盒子是十分倾侧的。

图 6 盒子|9 张[0]，[1]|的直方图。这直方图是十分倾侧的。



在抽取次数相当大之前，抽得数之和的概率直方图也是十分倾侧的，计算机已编制程序为从该盒中抽取 25、100 或者 400 次所得数之和作出概率直方图(见图 8)。抽取 25 次时，直方图在左端比曲线高许多，在右端比曲线低一些。正态近似不适用。抽取 100 次时，直方图对曲线的拟合好得多；抽取 400 次时，你必须仔细看才能发现差异。

再一个例子：考虑盒子|[1] [2] [9]|，盒中票的直方图在图 7 中给出。这张直方图看上去一点也不像正态曲线。

图 7 盒子|[1] [2] [9]|中所含票的直方图。该直方图十分不同于正态曲线。

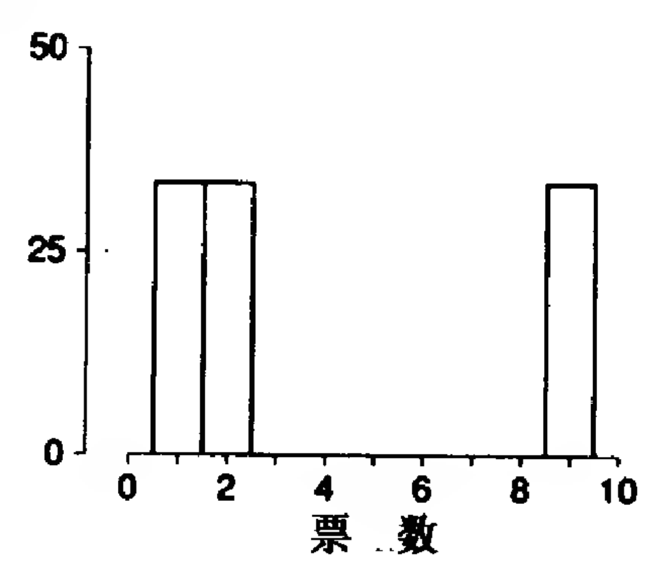
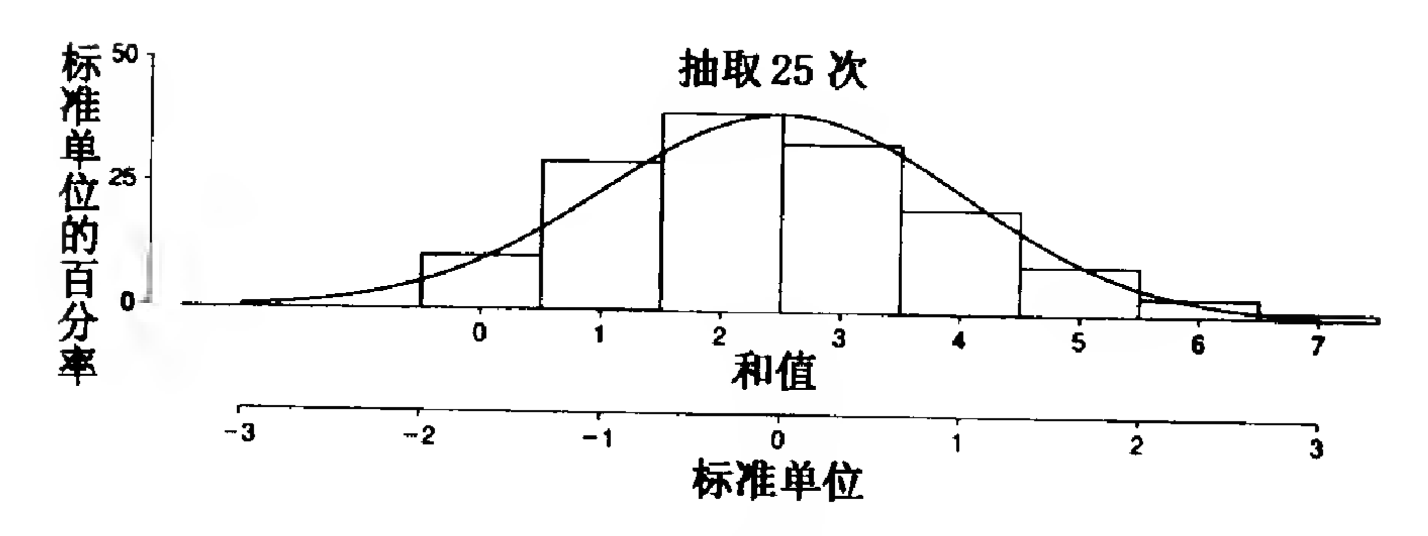
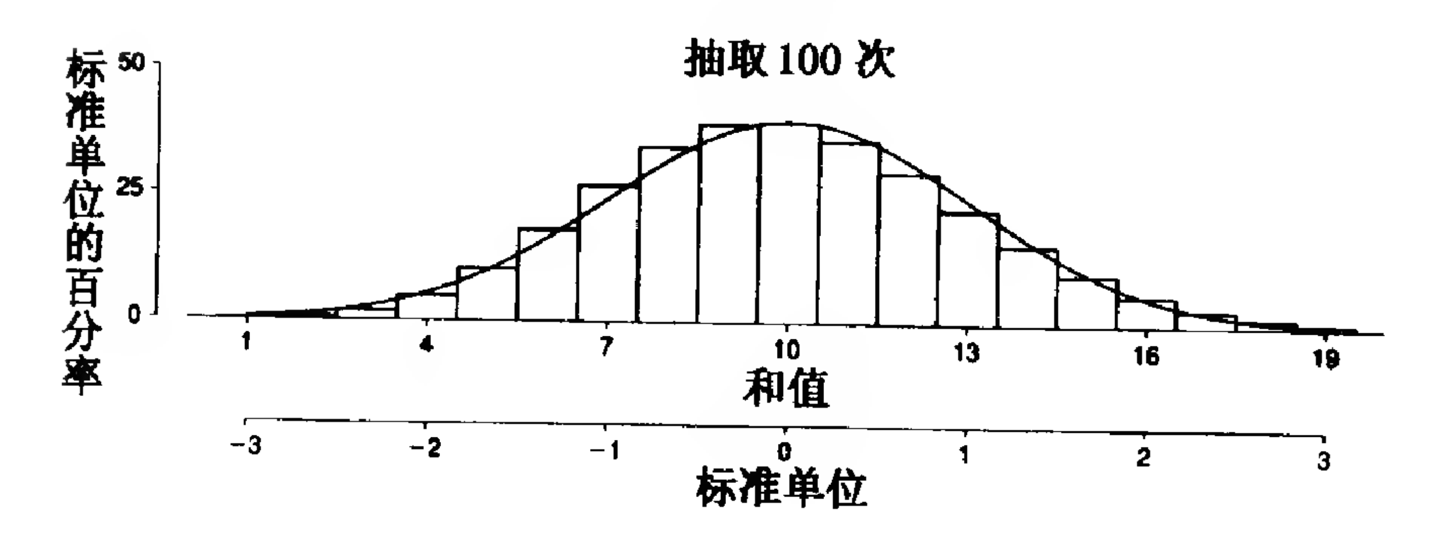
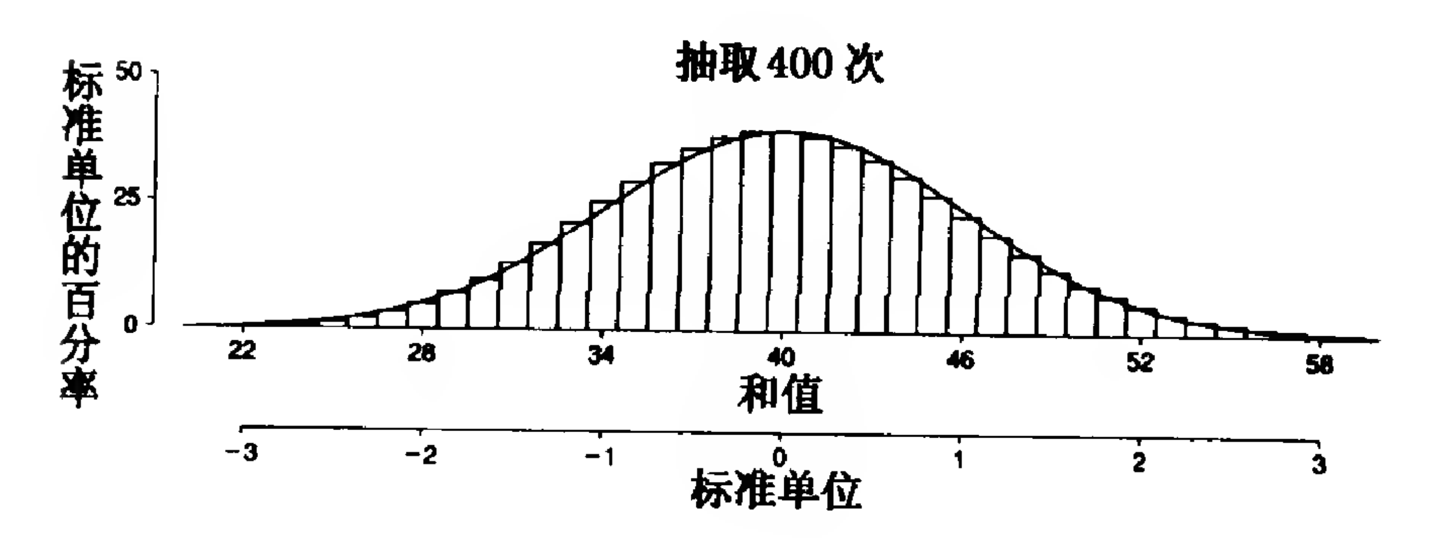


图 8，从盒子 9 张 $\boxed{0}$，$\boxed{1}$ 中抽得数之和的正态近似。顶部表示抽取 25 次所得数之和的概率直方图，中间是抽取 100 次的，底部是抽取 400 次的。标出正态曲线以作比较。这些直方图在左端比正态曲线较高，在右端较低，因为盒子是倾侧的[⑤]。当抽取次数增大时，直方图愈来愈紧密地遵循正态曲线。



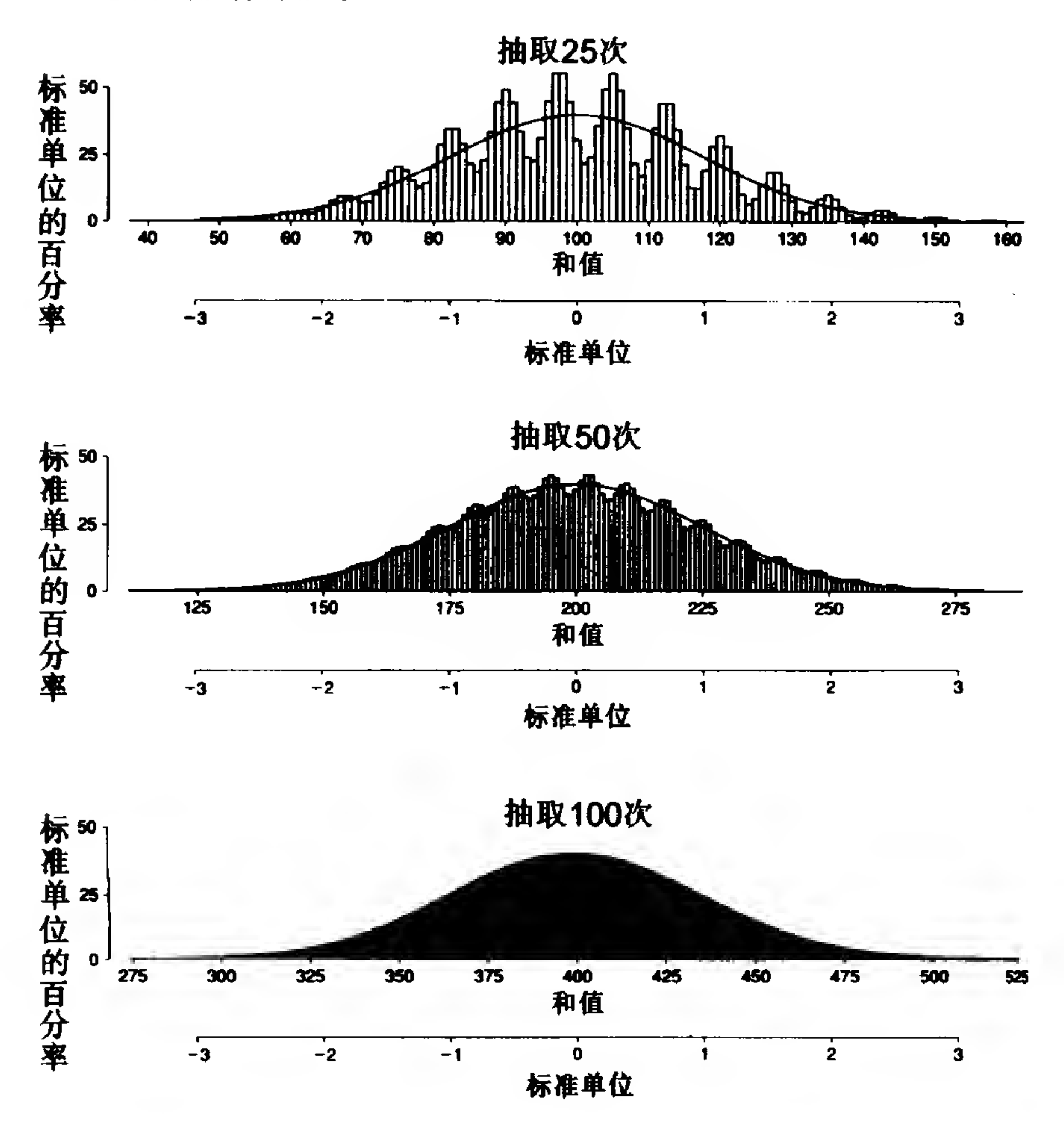




抽取 25 次时，和的概率直方图仍然与曲线十分不同

它显示波浪形（图 9）。抽取 50 次时，波浪仍然存在，但是小得多了。而抽取 100 次时，概率直方图与曲线已经没有什么区别了。

图 9 和的正态近似。根据从盒子 | 1 2 9 | 中抽取所得数之和绘制的直方图。顶部是抽取 25 次的，与正态曲线拟合得不很好（注意波浪[6]。）中间是抽取 50 次的。底部是抽取 100 次的，它与正态曲线拟合得很好。

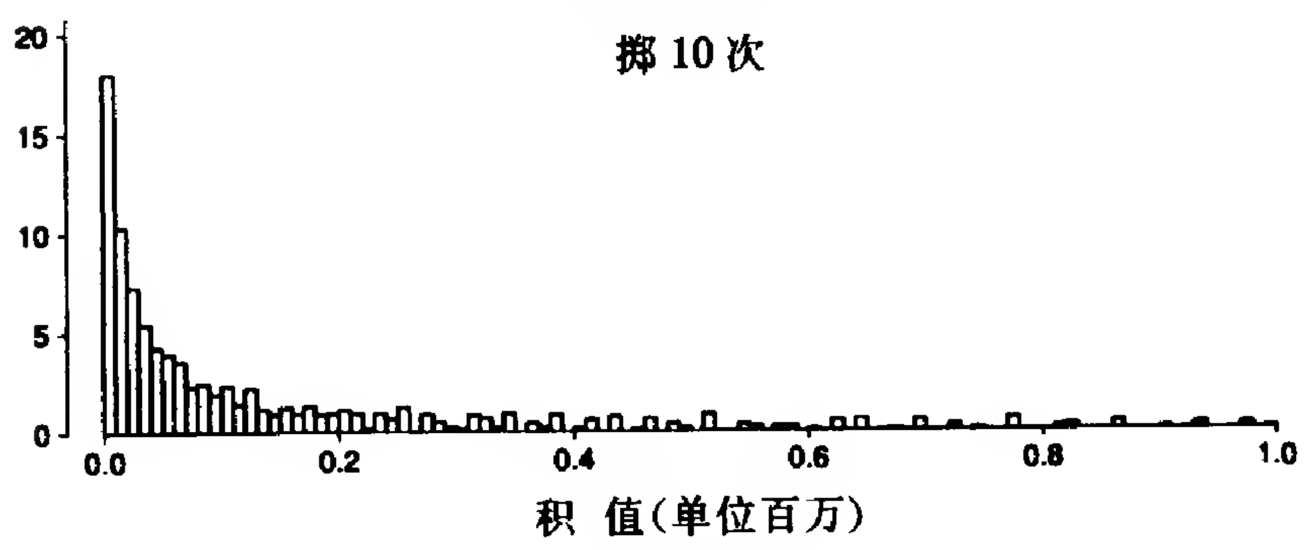


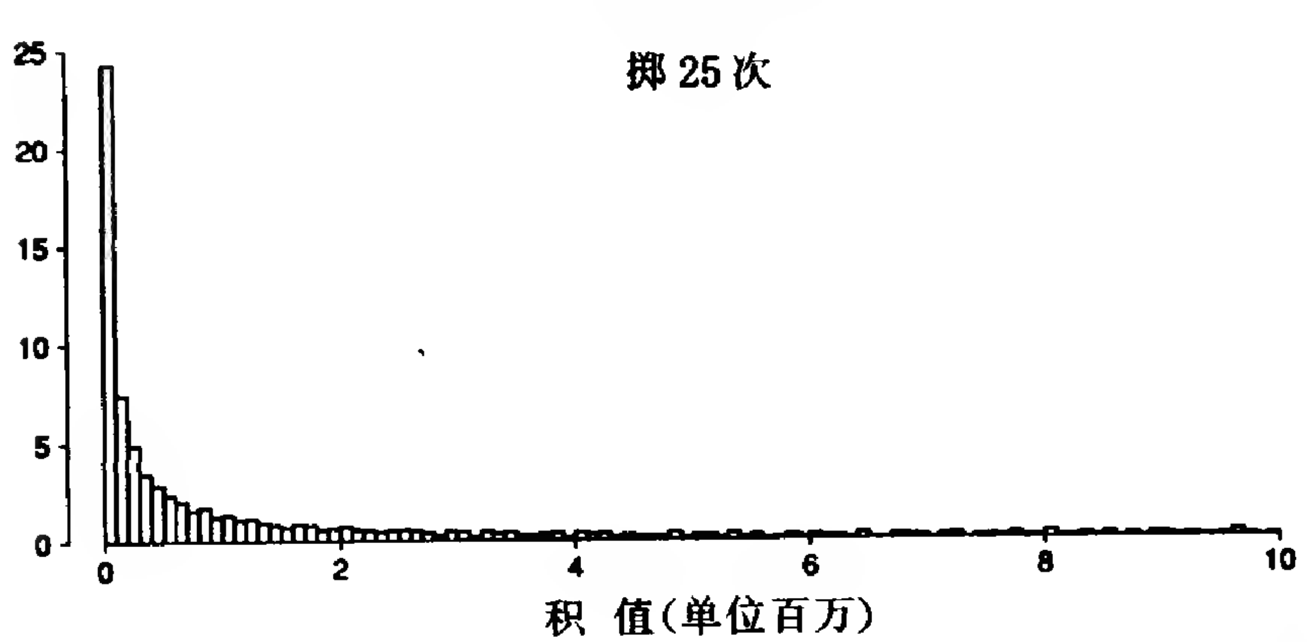
正态曲线与和相关联：例如，积的概率直方图往往是与正态曲线十分不同的。图 10 的顶部给出掷一颗骰子 10 次所得点数之积

的概率直方图。它一点也不象正态曲线。增大掷骰子的次数并没有使直方图更接近于正态：掷 25 次所得之积的概率直方图给出在底部，它甚至更糟⑦。乘法是与加法不同的。

对于掷 10 次，积的直方图显示的点至一百万；6%的面积超出这点，从而没绘出来。一百万看上去是一个很大的数，但是乘起来增加得很快。积的最大值是 6 自乘 10 次：$6^{10}=60466176$。与此相比，一百万毕竟就不那么大了。对于掷 25 次，积的最大可能值那真是一个大数：$6^{25}\approx3\times10^{19}$，即 3 后面跟 19 个 0。（相比之下，1989 年美国联邦债务“仅”3 万亿美元：亦即 $ 3 后面跟 12 个 0。）

图 10 一颗骰子掷 10 次和 25 次所得点数之积的概率直方图。这些直方图看上去一点也不像正态曲线。每个矩形的底覆盖积的一些值的区域；矩形的面积等于积取到这区域中某一值的机会，掷 10 次，约 6%的面积没绘出；掷 25 次，约 2%没绘出，在上部图中，纵向尺度是万分之一；下部图，则是 10^{11} 分之一。





习题 C

1. 一枚不均匀的硬币 10 次中有 1 次机会出现头像。将它抛 400 次。估计恰好得到 40 次头像的机会。

2. 将习题 1 中的硬币抛 25 次，假定用正态近似来估计恰好得到 1 次头像的机会，估计是否大体上令人满意？还是太高？太低？不必计算：只要看图 8。

3. 同一枚硬币抛 100 次，如果要求你估计得到 10 次或少于 10 次头像的机会，你必须留意矩形的边界吗？不必计算，看图 8 即可。

4. 本习题参照图 9 顶部图，它表示从盒子 | 1 2 9 | 中抽取 25 次所得数之和的概率直方图，和是 100 的机会等于

   (i)在 99.5 与 100.5 之间概率直方图下的面积

   (ii)在 99.5 与 100.5 之间正态曲线下的面积

   选择一种，并说明为什么？

5. 从盒子 | 1 2 9 | 中抽取 25 次所得数之和最大可能等于____，最小可能等于______，即使它的期望值是______。请从

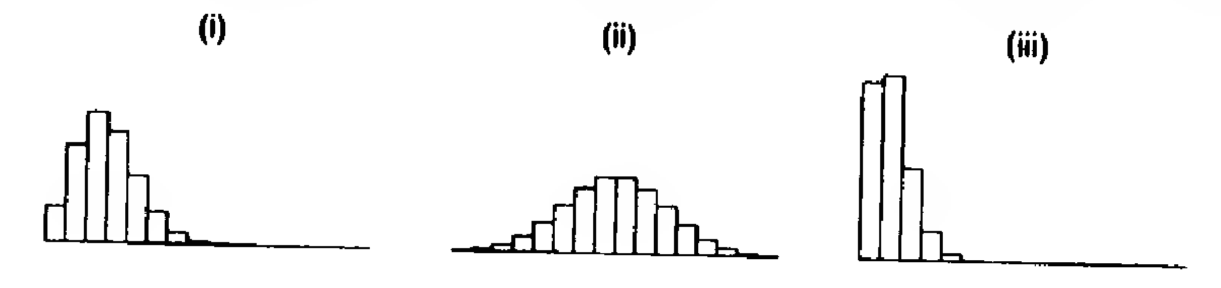
   100　　101　　102　　103　　104　　105

   中选择填空，不必计算，看图 9 即可。

6. 随机放回地从下面的盒子中各抽取 25 次：



   下面所给出的和的概率直方图次序被打乱，试将直方图与盒子配对。



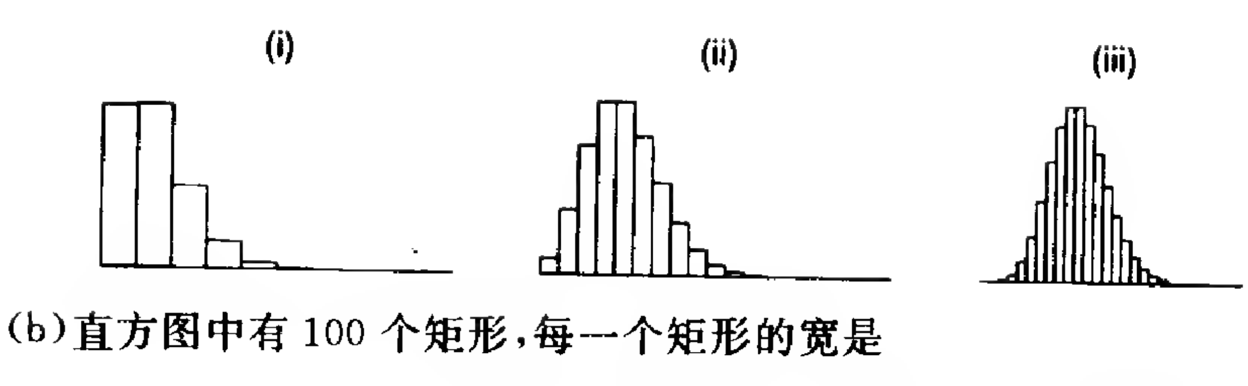
7. 下面所给出的是从盒子 | 99 个 0 1 | 中分别抽取 100、400 和 900 次所得数之和的概率直方图，哪一个直方图是哪一种抽取次数的？（见下页图）

8. 本题参照图 10 中顶部图。

   (a)积的期望值接近 276 000，积超过这个数的机会是

   差不多 50%　　远大于 50%　　远小于 50%

   选择其中之一，并说明之



(b)直方图中有 100 个矩形，每一个矩形的宽是

1　　10　　100　　1 000　　10 000　　100 000　　1 000 000

选择其中之一。

(c)哪一个较可能是积的范围？

390 000—400 000　　400 000—410 000

这些习题的答案在第 707—708 页上。

6. 结论

本章的要点可以陈述如下

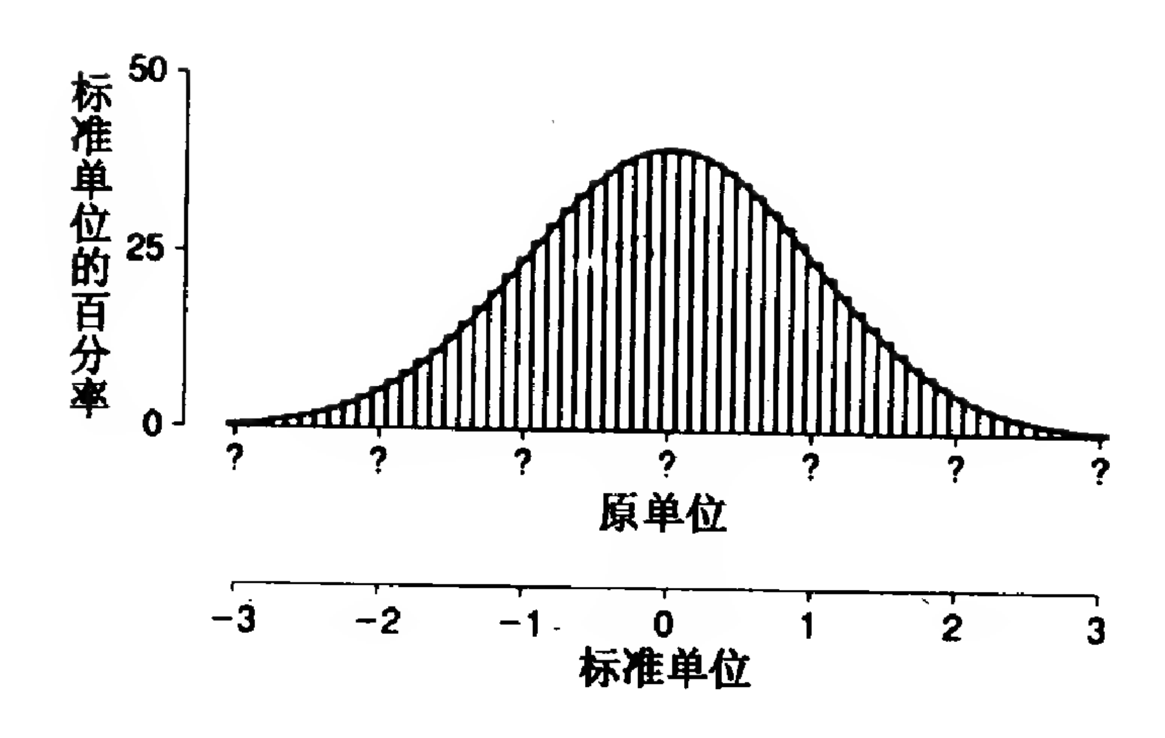
> 当随机放回地从一个盒子中作抽取时，即使该盒子中所装票子并不遵循正态但抽得数的和的概率直方图将遵循正态曲线。直方图必须换算成标准单位，并且抽取的次数必须适当地大[8]。

这个事实是本书余下部分中统计方法的关键。

抽多少次才够呢？没有现成的答案，因为很多地依赖于盒子的内含（别忘了图 9 中的那些波浪），但是，对许多盒子，就大多数用途来说，抽取 100 次所得数之和的直方图与正态曲线是足够接近了。

当概率直方图拟合于正态曲线时，它可以用期望值与标准误差来刻划。例如，假如你必须在没有进一步信息下画出这样一张直方图，以标准单位，你至少能够为它作出第一次近似：

要完成此图，你必须将图中的问号填写明白而把标准单位化回原单位。这就是期望值与标准误差所做的。因为它的形状恰好与正态曲线一致，因此关于这直方图你想知道的一切，它们几乎都能解释。



期望值把概率直方图的中心定在水平轴上，而标准误差确定了它的散布。

使用平方根法则时，和的期望值与标准误差可由

- 抽取次数
- 盒平均数
- 盒的 SD

算出。因此这三个量几乎决定了和的行为，此即为何盒子的 SD 成为其散布程度的一个重要度量的理由[⑨]。

本章讨论了两类直方图的收敛性，重要的是区分它们。在图 1 中，从盒子 [1] [2] [3] [4] [5] [6] 中抽取的次数是固定的：为 2。基本机会过程是从盒子中抽两个数并求和。这一过程重复的次数愈来愈多：100，1 000，10 000。和的观察值的经验直方图（数据直方图）收敛于概率直方图（机会直方图）。在第 5 节中，从盒子中抽取的次数愈来愈大。于是和的概率直方图愈来愈光滑，其极限变成正态曲线。

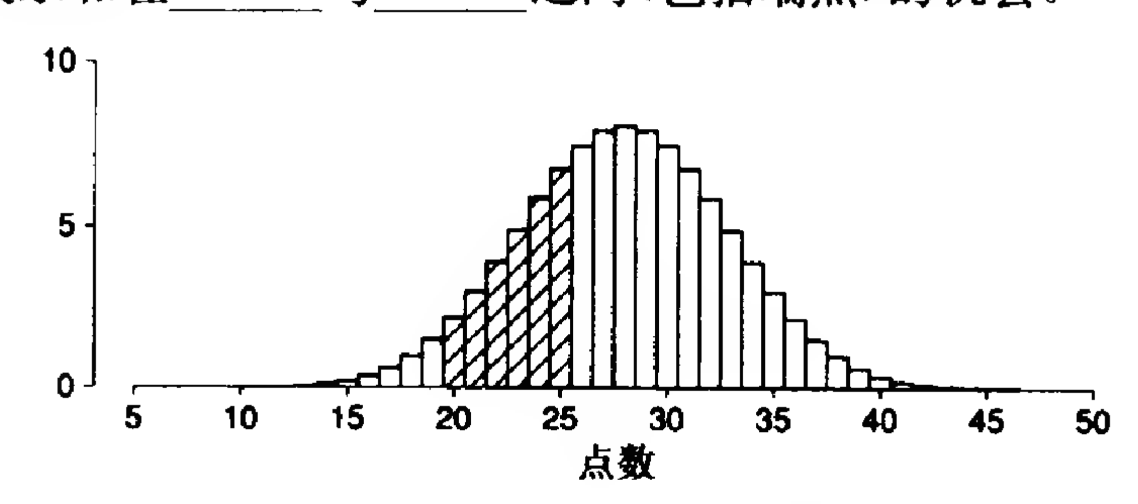
在本书的第二部分中，将正态曲线用到了数据上，在某些情况，可以用本章所讨论的两种收敛从数学上论证这样做的正确性。当重复次数大时，经验直方图将接近于概率直方图，当抽取次数大

时，和的概率直方图将接近于正态曲线。因此，当重复次数和抽取次数都大时，和的经验直方图将接近于正态曲线[10]。这完全是一个纯粹的逻辑推理问题；数学家能证明每一步。

但是在解释为什么一张数据直方图会看上去像正态曲线时，还有一些事被遗漏了。必须证明，产生数据的过程类同于从一盒子中抽取数字并求其和。这类论证将在第七部分中讨论，并且涉及的不止数学。届时同样存在论据问题需要解决。

7. 复习题

1. 下图表示一颗骰子掷 8 次所得点数之和的概率直方图。阴影面积表示和在______与______之间（包括端点）的机会。



2. 随机放回地从盒子 [1] [3] [5] [7] 中抽取 400 次。

   (a)估计抽得数之和大于 1500 的机会

   (b)估计少于 90 张 [3] 的机会

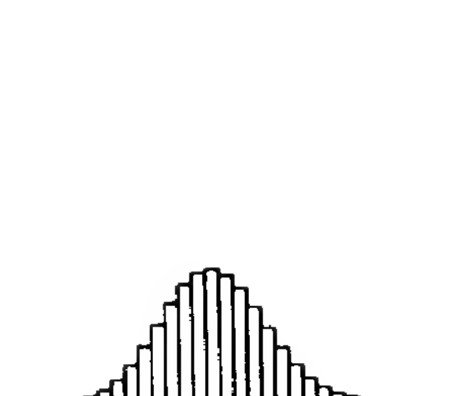
3. 随机放回地从盒子 [0] [1] [2] [3] 中抽取 10 次。和将在 10 到 20（包括端点）区间内的机会等于______与______之间______下的面积。填充空白。对于最后一个空白，供你选择的是：正态曲线、和的概率直方图。解释你的答案。

4. 一枚硬币抛 25 次。估计得到 12 次头像和 13 次背面的机会。

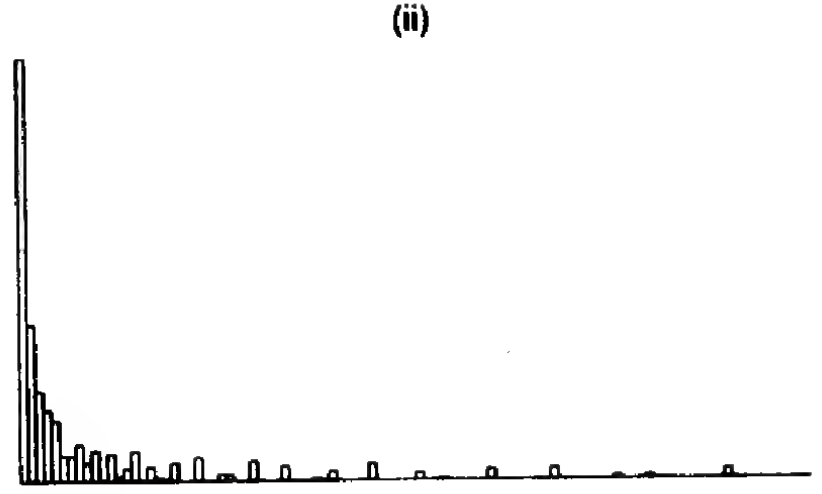
5. 随机放回地从盒子 [1] [1] [2] [2] [3] 中抽取 25 次，下面图中一张是抽得数的直方图，一张是抽得数之和的概率直方图，还有一张是积的概率直方图。哪一张图是哪一种情况，为什

么?

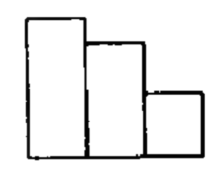
(i)



(ii)



(iii)



6. 一位程序编制者根据新程序 TOSSER 模拟抛一枚硬币。作一次初步检查,他指令程序抛一百万次。这程序回答的计数是 502 015 次头像。这位程序编制者看了这个计数后想:

   唔,差 2 015,够多的。不,等一下。把它与一百万作比较。不计零头 15,一百万次中的 2 000 次是 1 000 次中的 2 次。那是 500 分之一。一个百分数的五分之一,十分小。好,TOSSER 行。

   这个推理合理吗?还是不合理?请为你的选择辩护。

7. 一盒装有 10 张票,4 张标上正数,6 张标上负数。所有的数都在 −10 与 10 之间,随机放回地从该盒中抽取 1 000 次,要你估计抽得数之和是正的机会。

   (a)在已经给出的信息基础上你能做这件事吗?

   (b)如果另外再告诉你盒中数的平均数与 SD,但不告诉你这些数本身,你能做这件事吗?

   简短地说明之。

8. 重复习题 7，假如要你估计得到不少于 425 个正数的机会。

9. 重复习题 7，假如要你估计得到不少于 100 张[3]的机会。

10. 随机放回地从一个盒中抽取，而且抽取次数愈来愈大。说出下面每一陈述是正确的，还是不正确的，并加以说明。

(a)和的概率直方图(以标准单位)与正态曲线拟合得愈来愈紧密。

(b)盒中所装票子的直方图(以标准单位)与正态曲线拟合得愈来愈紧密。

(c)抽得数的直方图(以标准单位)与正态曲线拟合得愈来愈紧密。

(d)积的概率直方图(以标准单位)与正态曲线拟合得愈来愈紧密。

11. 抛一对骰子，其点数之和如同

(i)从盒子|[2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]|

中作一次抽取

(ii)从盒子|[1] [2] [3] [4] [5] [6]|

中抽取两次所得数之和。

两者选一，并说明之

12. 随机放回地从盒子|[1] [2] [9]|中抽取 25 次。

(a)一位统计学家用正态曲线去计算抽得数之和等于 90 的机会。结果是

太低　　太高　　大约正确

选择一种，并说明之

(b)对于和在 90 与 100 之间的机会，再解题

不必计算；看图 9 即可

8. 小结

1. 当获得一个和的机会过程重复许多次时，观察值的经验直

方图趋于概率直方图。

2. 概率直方图用面积来表示机会。

3. 随机放回地从一盒中抽取时，和的概率直方图将遵循正态曲线，尽管盒中所装的票并不遵循正态曲线，直方图必须置于标准单位下，而且抽取次数必须适当地大。

4. 正态近似是在计算面积前用正态曲线代替实际概率直方图。通常可以通过留意矩形的边线来改善近似的精度。

5. 遵循正态曲线的概率直方图能够用期望值和 SE 十分好地刻划。期望值确定概率直方图的中心，而 SE 度量其散布程度。

第六部分　抽　　样

19
抽样调查

"数据！数据！数据！"他不耐烦地喊道，"我不能做无米之炊。"

Sherlock Holmes[①]

1. 引言

调查人员通常需要归纳整个的一类个体。这个类称为总体。譬如，在预测美国总统选举的结果中，一个贴切的总体是全体合法选民组成的类。但是，研究整个总体一般并不实际。能检测的只是它的一部分，这部分称为样本。调查人员将根据部分对整体做出归纳；更为专业的措词是，他们根据样本对总体做出推断[②]。

通常存在着调查人员需要知道的关于总体的某些数值特征。这些数值特征称为参数。在预测美国的总统选举中，有关的两个参数是：

- 全体合法选民的平均年龄。
- 当前登记投票的全体合法选民的百分数。

一般情况下，象这样的一些参数是不能精确测定的，仅能根据样本估计。于是，精度成了主要问题：估计将会达到怎样的接近程度？

参数由统计量或可根据样本算得的某些数值估计。譬如，用10 000 名美国人组成的样本，为了估计上述两个参数：调查人员可以计算下面两个统计量：

- 样本中合法选民的平均年龄。
- 样本中当前登记投票的合法选民的百分数。

统计量是调查人员知道的，参数是他们需要知道的。

只有当样本代表了总体时，根据样本估计参数才合理。要核对这一点只查看样本是不可能的。道理很简单。要了解样本和总体在一些关键方面是否相似，调查人员必须先知道他们所要估计的总体的有关数值特征——一个恶性的循环。因此，你必须查看样本是如何选取的。有些选取样本的方法效果较差，有的很可能提供有代表性的样本。本章的两个主要课目是：

- 选取样本的方法至关重要。
- 最好的方法包含有计划地引用机会。

（在试验设计中，当指派试验对象去处理组或对照组时也会出现类似的问题：见第一部分。）

2. The Literary Digest《文学摘要》民意测验

1936 年，Franklin Delano Roosevelt（富兰克林·迪拉诺·罗斯福）任美国总统的第一任期届满。这年是选举年，共和党的候选人是 Kansas（堪萨斯）州州长 Alfred Landon（艾尔弗雷德·兰登）。国家正努力从大萧条中恢复过来，失业人数仍高达九百万人。在 1929－1933 这段期间实际收入下降了约 1/3，并且刚开始回升。Landon 正指责政府的一项经济计划，而罗斯福在为他的财政赤字辩解[3]。

Landon.　　挥霍浪费的人必须离任。

Roosevelt.　　在我们能够平衡联邦政府的预算之前，必须先平衡

美国人民的预算。这已是常识，不是吗？

纳粹分子正重新武装德国，而西班牙的内战也正推到绝望的高峰。这些问题和争端占据了《纽约时报》的头条新闻，但是都被两位候选人忽视了。

Landon.　　我们应该专心致力于自己的事务。

绝大多数观察家认为罗斯福会是不费力的获胜者。《文学摘要》杂志却不这样认为，它预测 Landon 会以 57%对 43%的压倒优势获胜。这预测是根据有约二百四十万人参加的一次民意测验作出的。《摘要》的显赫威望支持着它的预测，自 1916 年以来的历届总统选举中它都正确地预测出获胜的一方。但是，Roosevelt 以 62%对 38%的一边倒优势赢得了 1936 年的选举。（此后不久《文学摘要》就垮了。）

《摘要》的预测误差幅度之大令人吃惊。这是重要民意测验曾作出过的最大误差。这么大的误差是怎么得来的呢？接受民意测验的人数大得绰绰有余。事实上，George Gallup（乔治·盖洛普）才刚刚设立起他的调查机构[4]。用他的方法，他能在《摘要》公布其预测之前以仅一个百分数点的误差预言《摘要》的预测结果，利用另一个约 50 000 人的样本，他正确地预测了 Roosevelt 将获胜，尽管他对 Roosevelt 所得选票的预测稍微偏离些。盖洛普预测 56%赞成 Roosevelt，而实际百分数是 62%，误差为 62%－56%＝6 个百分数点。（调查机构用“百分数点”作为实际与预测百分率之间差的单位。）各种结果汇总在表 1 中。

表 1 1936 年的选举

	Roosevelt 的百分数
盖洛普预言《摘要》的预测结果	44
《摘要》预测的选举结果	43
盖洛普预测的选举结果	56
选举结果	62

注:上述百分数仅用主要政党所得选票计算,选举中约有 2%的选票投向小党的候选人。

摘自:George. Gallup:The Sophisticated Poll—Watcher's Guide.(超级民意测验——观夜者指南)1972。

若想找出《摘要》在哪一点上出了差错,你必须查一下他们是怎样抽取样本的。抽样的程序应该合理,以公平方式选择人选作为样本中的成员,以便获得公众的一个有代表性的横剖面。抽样程序将这一类或那一类人排除在样本之外所表现出的系统倾向称为选择偏性。《摘要》的程序是将问卷邮寄给一千万人,这一千万人的名字和地址摘自诸如电话簿或俱乐部会员名册。这导致筛选掉那些不属俱乐部成员或没有安装电话的穷人。(譬如,当时四个家庭中仅有一家装电话。)因此,在《摘要》的抽样程序中存在排挤穷人的很强的选择偏性,在 1936 年之前,这种偏性可能对预测影响不大,因为富人与穷人以类似的思考投票。但在 1936 年,政治上的划分更紧密地遵循经济路线:绝大多数穷人投 Roosevelt 的票,而富人赞成 Landon。《摘要》的预测误差如此之大的原因之一是选择的偏性。

当选择程序有偏时,抽取一个大的样本并无帮助。它只不过是在较大的规模下去重复基本错误。

因此《摘要》在抽样的第一步做得很糟。还有第二步。在确定那些人应选入样本之后,调查机构还需要去获得他们的想法。这比想像的要难些。若被选入样本的大多数人实际上不回答问卷或询问,那将产生严重的扭曲。称为不回答偏倚。

不回答者区别于回答者的明显之处:他们没有回答。经验表明

他们在其它重要方面也同样存在差异[5]。譬如,《摘要》在1936年做过一次特别调查,给在芝加哥登记的选民每三个人寄去一份问卷。约20%回答,其中超过半数赞成Landon。但是选举中芝加哥以二对一的差幅赞成Roosevelt。

> 不回答者可能非常有别于回答者,当出现高不回答率时,谨防不回答偏倚。

在最重要的《摘要》民意测验中,收到问卷的一千万人中只有二百四十万人费心回答。这二百四十万人甚至还代表不了被测验的一千万人,更不用说全体选民总体。《摘要》民意测验被选择偏倚和不回答偏倚两者搞糟了[6]。

为了度量回答者与不回答者之间的差异,当时做了一些特别调查。结果表明低收入和高收入的人倾向于不回答问卷,因此中等收入阶层在回答者中的代表性超出比例。由于上述原因,现代调查机构更喜欢采用亲自询问来代替邮寄问卷,亲自询问的典型回答率是65%,而邮寄问卷仅25%[7]。尽管如此,亲自询问仍然存在不回答偏倚问题。访问员来访时,不在家的人与在家接受访问的人可能在工作时间,家庭关系,社会背景等方面有较大的差别,从而看法上也就不一样。好的调查机构记住了这问题,且有巧妙的处理方法。(见第六节)

> 某些样本确实很差,若想了解一个样本是否可取,查查它是怎样选取的。有选择偏倚吗? 有不回答偏倚吗?

再回到1936年的选举,Gallup是怎样预测出《摘要》的预测的?他只从《摘要》要用的名单中随机选出3 000人,并给他们每人寄去一张明信片询问他们打算怎样投票,他明白随机样本很可能具有代表性,就象下面两章将介绍的那样。

3. 民意测验Dewey(杜威)当选年

Thomas Dewey(托马斯·杜威)在纽约市,以一名嫉恶扬善

的地方检查官成名，继而入主在 Albany（奥尔班尼）的州长官邸。1948 年他是共和党的总统侯选人，挑战在位的 Harry Truman（哈里·杜鲁门）。Truman 在 Kansas 市（堪萨斯）以 Boss Pendergent 的门生开始政治生涯。选进参议院后，Truman 出任联邦政府副总统，Roosevelt 逝世他接任总统职务。Truman 是本世纪最有效率的总统之一，也是最富色彩的总统之一，在他的办公桌上，他总放着一则告示，“切莫推卸职责”。他特别喜爱的另一格言成了美国政治术语的一部分：“如果你受不了热，就呆在厨房外面。”但是在 1948 年 Truman 处于劣势，因为那是一个动乱的年代，第二次世界大战勉强结束，冷战状况下令人焦虑的半保障和平刚开始，国内不太平，国外情况错综复杂。

三家主要民意测验采访了竞选情况：代表 Hearst（赫斯特）报的 Crossley（克劳斯莱）；联合了全国约 100 家独立报纸的 Gallup；以及代表 Fortune（《幸运》）杂志的 Roper（罗伯尔）。到了秋天，三家都声称 Dewey 是获胜者，且将领先约 5 个百分数点。Gallup 的预测基于 50 000 人次的询访；而 Roper 是 15 000 人次。正如 Scranton Tribune 所表述的：

统计资料使 Roper 相信 Dewey 实际已当选。

但是在选举日，Truman 以稍低于 50%的公众选票获得逆转性胜利；Dewey 得票仅略高于 45%。选举的结果列于表 2。

表 2 1948 年的选举

侯选人	预测			结果
	Crossley	Gallup	Roper	
Truman	45	44	38	50
Dewey	50	50	53	45
Thurmond	2	2	5	3
Wallace	3	4	4	2

来源：F. Mosteller 等，1948 年的预选民意测验。（纽约：社会科学研讨会，1949）

为找出上述民意测验出了什么差错，必须查出他们是怎样选取他们的样本的[8]。他们全都采用所谓的定额抽样方法。根据这一

程序，访问人员访问的各种对象都规定了固定的定额；此外某些范畴（如居住地区，性别，年龄，种族和经济状况）的人数也是固定的。在其它方面，访问人员可以自由挑选他们喜欢的对象。举个例子，Gallup 民意测验要求在 St. Louis（圣·路易斯）的访问人员访问 13 个对象，并规定其中[9]：

- 6 人住在近郊，7 人住在市中心。
- 男的 7 人，女的 6 人。

规定 7 个男的（对女的也有类似的定额规定）中，

- 3 人 40 岁以下，4 人 40 岁以上。
- 1 名黑人，6 名白人。

6 名白人支付的月租又做了以下规定，

- 1 人的支付金额不少于 44.01 美元
- 3 人的支付金额为 18.01 美元～44.00 美元。
- 2 人的支付金额为不超过 18 美元。

提请注意，这里指 1948 年的价格。

从常识观点看，定额抽样是好的。似乎它保证了样本和选举总体在被认为对选举行为有影响的所有主要特征方面将会相似。（住处，性别，年龄，种族和租金的分布可由普查局的数据十分接近地估计。）但是 1948 年的经验表明这种程序的效果很差，现在我们分析一下原因。

调查机构想要一个能忠实代表国民政治见解的样本，然而，对共和党和民主党的得票数是不可能安排定额的。因为国民中政治见解的分布状况恰恰是调查机构所不知道以及正努力尝试去发现的。因此，对其它变量的各种定额只不过是企图让样本反映国民政治的一种间接的努力。可是除了调查机构加以控制的因子之外，还有影响选举行为的其它许多因子。住在近郊的富有的白人男子中有投民主党票的，住在市中心的贫苦黑人妇女中有投共和党票的。结果是：对于调查机构，完全可能出现精心选出的样本在所有的人口统计变量方面是国民的理想的横剖面，但是过后却发现样本的

投票是一个样子而国民的投票又是另一个样子。在 1948 年之前，这种可能性谅必象是理论上的。

反对定额抽样的下一个理由是非常重要的。因为，它包含了方法上的一个至关重要的特征，且初次使用时常常不容易觉察：在规定的定额内，访问人员可以自由选择他喜欢的任何人。这给人为选择留有过多的余地。而人为选择常易带偏性。1948 年，访问人员选择了过多的共和党人。总的说来，共和党人比民主党人较为富裕并受过较好的教育。他们更可能拥有自己的电话和有永久性住址，并居住在较好的街区。在每个人口统计大组中，共和党人相对较易访问。如果你是一个访问人员，或许你也会最终挑选过多的共和党人。

事实上，从 1936 年到 1948 年的历届总统选举中，访问人员喜欢采访共和党人，情况正如表 3 列举的 Gallup 民意测验结果所示。在 1948 年之前，民主党领导的政绩是那么显赫，因此克服了民意测验中的共和党偏性，从而 Gallup 能预测获胜者。在 1948 年，定额抽样中的共和党偏性压倒了大不如昔的民主党的领先。

> 在定额抽样中，样本被精心挑选以使在某些关键特征上与总体相似。这方法似乎合理，但并不怎么奏效。原因是无意的偏倚。

表 3 Gallup 民意测验中的共和党偏性，1936—1948。

年份	Gallup 预测的共和党得票百分数	共和党实际得票百分数	有利于共和党的误差
1936	44	38	6
1940	48	45	3
1944	48	46	2
1948	50	45	5

注：除 1948 年外，百分数都只根据主要政党得票数计算。

来源：F. Mosteller 等，1948 年预选民意测验（纽约：社会科学研讨会，1949）。

定额抽样中的定额是相当符合实际的，尽管它们不能保证成

功——差得远。但是真正灾难性的是访问人员自由挑选配与定额的方法[10]。可供选择的办法是用客观且不带偏袒的机会手段挑选样本。这是下一节的话题。

4. 调查工作中使用机遇

即使在 1948 年，一些调查机构已使用了机会，或概率方法去选取他们的样本。自 1948 年以来，事实上所有的调查机构都使用机会方法。机会是怎样用于抽取样本的呢？仅作开始，设想要在一个有 1 000 名合法选民的小镇上落实一项有 100 名选民参予的民意调查。那么，列出全体合法选民是可行的，将他们每个人的名字分别写在票子上，然后将 1 000 张票子全放入一只盒子中，并随机地抽出 100 张。

由于不存在对同一个人做二次访问的条款，因此抽样是不放回的。换句话说，先摇匀盒中的票，然后从中随机抽出一张放在一边，留下 999 张在盒中。再摇动盒子，抽出第二张并放在一边。反复进行直到抽足 100 张为止。

被抽到名字的人构成样本。这种样本称为简单随机样本。名字是被随机不放回地抽出。在每次抽取时，盒子中的每个名字有相同的机会被选中。访问人员无权决定他们的访问对象，这种方法不含偏袒性：每个人都有同样的机会入选样本。因而，平均数律保证了样本中民主党人的百分比很可能地接近于总体中的百分比。

简单随机抽样意即随机不放回地抽取。

对更为实际的场合，譬如 Gallup 民意测验试图预测总统选举，会出现什么情况呢？一种意见是取一个数千名合法选民的全国性简单随机样本。与定额抽样相比较，这将产生好得多的结果。但是，它绝不象听起来那么容易。在统计意义下随机抽出名字是件艰巨的工作。它和任意挑选人毫无相同之处。

要开始随机抽取合法选民，你需要有一张全体合法选民的表——近二亿个名字。首先是没有这样的一份表。即使有，要从二亿

个名字中随机抽出数千个本身就是难办的事。(别忘了,每次抽的时候盒子中的每个名字必须有同等机会被选中。)即使能抽出一个简单随机样本,抽出的人可能会分散在全国各地。派遣访问人员到各地找到他们的费用将会高得难于承受。

抽取简单随机样本恰恰是不实际的。因此,绝大多数调查机构采用多阶分群抽样的方案。名称是够复杂的,详情亦如此。但想法却是简单明确的。这将在 1952 至 1984 这期间 Gallup 预选调查的有关部分里描述;这些调查基本上是用相同的方法进行。Gallup 民意测验在美国的四个地理区域—— 东北,南,中西,西(图 1)中的每一个展开独立的研究。在每一区域内,他们把规模差不多的居民中心区归成群,某这样的群可以是东北地区所有的人口在 5 万到 25 万之间的城镇全体。然后再从这些城镇中选出一个随机样本。派遣访问员到这些选出的城镇,对群中其它的城镇则不做访问。其它群也以相同方式处理。这完成了抽样的第一阶段[11]。

为了选举的效果,每个城镇都划分成选区,而选区再分成选举分区。在抽样的第二阶段,从前阶段选出的每个样本城镇随机选出若干选区。在抽样的第三阶段,从前面选出的每一选区中随机抽出若干选举分区。在第四阶段,从每一选出的选举分区随机抽出若干家庭[12]。最后是访问选出的家庭中的某些成员。就是到了这一步访问人员也不容许有决定权,譬如,Gallup 民意测验的访问员会接到通知"与家庭中 18 岁以上的最年青的男士交谈,若没有男子在家,则与 18 岁以上最年长的女士交谈。"[13]

这种设计安排提供定额抽样的许多长处。譬如,提供了按住处样本的分布与全国住户的分布相似。但是,它在挑选过程中的每一阶段都采用客观且不带偏袒性的机会手段来选取样本的单元,这完全消除了定额抽样的最坏的特征:访问员方面的选择偏性。

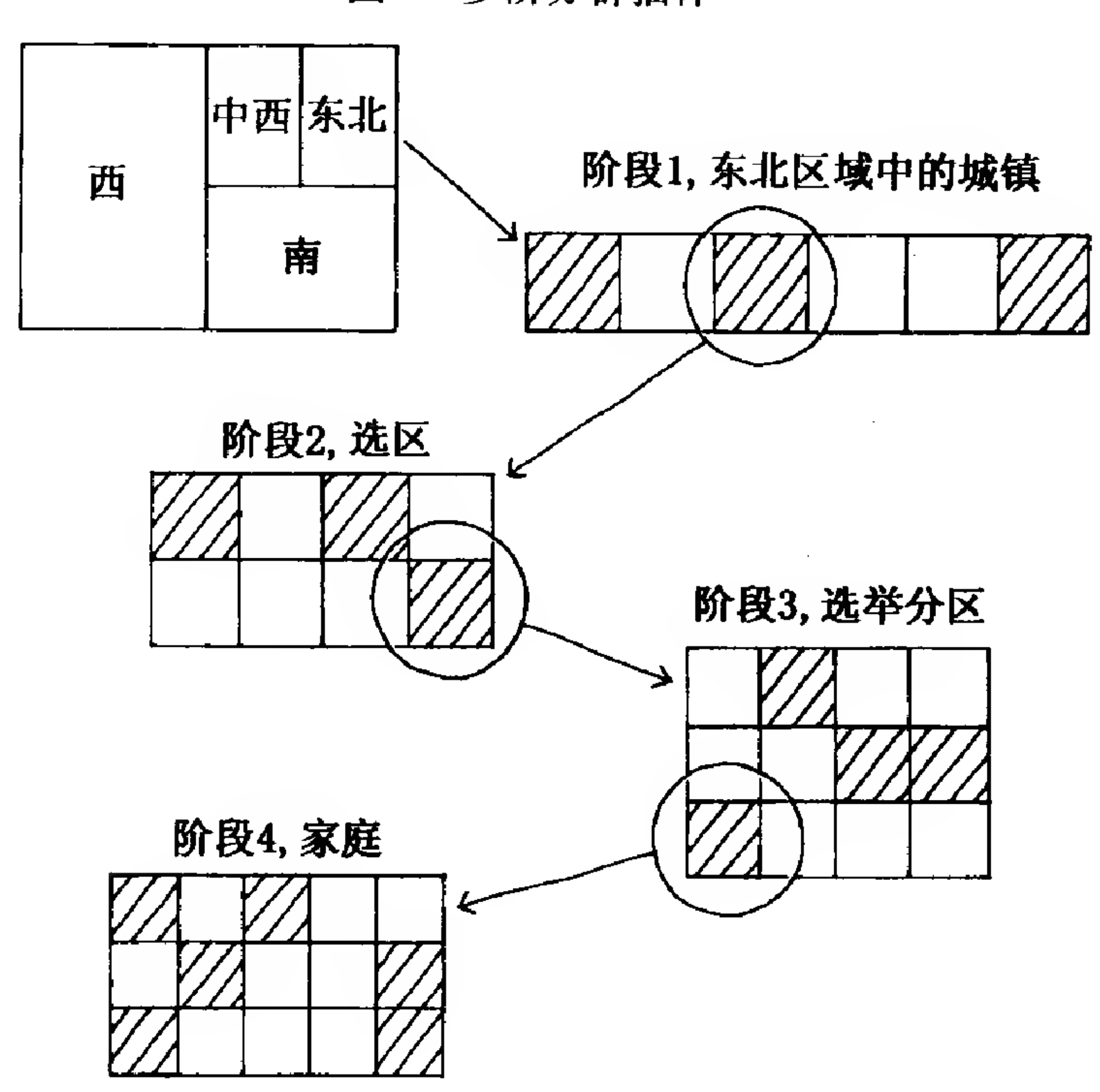
简单随机抽样是基本的概率方法。其它方法可能相当复杂,但是抽样的所有概率方法都有下面两个重要特征:

- 访问员对访问谁毫无决定权。

· 选样本有明确的程序，且包含机会的有计划使用。

因此，用概率方法有可能计算出总体中任一特定个体被选入样本的机会[14]。

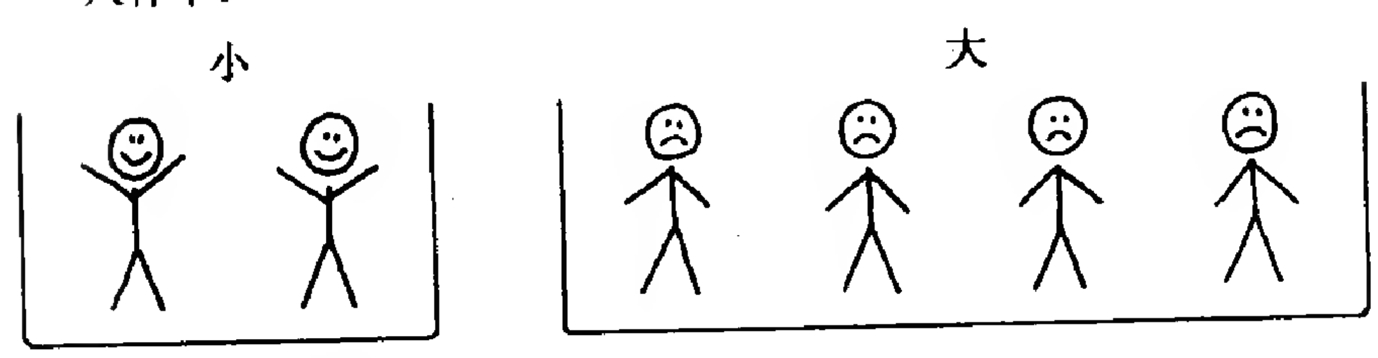
图 1　多阶分群抽样



定额抽样不是一种概率方法：它通不过上述两项检验标准。访问员在选择访问对象上有太多决定权。机会仅以一种非常盲目和无计划的方式参于其中：访问员喜欢接近哪种类型的人？一天中的某一特定时刻谁会在某一特定街上行走？没有一个调查机构能对这类机会做出数值估计。

通常，概率方法都设计成总体中的每一个体都有相同的机会进入样本。然而，Gallup 的方法在某些次要方面是有偏性的，它不利于生活在大家庭中的人（见图 2）。因此，作了某种调整以纠正这一偏性（见第 6 节）。

图 2　家庭的偏性：从每一家庭中挑选一人会产生一种不利于生活在大家庭中的人的偏性。设想从如下两个家庭中随机挑选一人：然后从选出的家庭中随机挑选一人，它产生一个容量为 1 的样本。在小家庭中的成员比在大家庭中的成员有较高的机会被选入样本。



5. 概率方法奏效有多大?

自 1948 年以来，Gallup 民意测验和所有其它重要民意测验都用概率方法选取他们的样本。盖洛普民意测验在 1948 年后的总统选举中的记录列于表 4。有三点值得注意。样本的容量急剧下降：在 1948 年，盖洛普民意测验使用容量约为 50 000 人的样本，而现在他们用这容量的五分之一到十分之一的样本。不再有偏向共和党或民主党的任何一贯倾向。并且精度明显提高。从 1936 年到 1948 年，误差约为 5%；在此之后，它们要小得多。用概率方法挑选样本，从 100 000 人中只取不满 5 人，Gallup 民意测验已能以惊人的精度预测选举。这记录证实了抽样中概率方法的价值。

为什么机会方法会如此奏效呢?起先，似乎认为判断对选择样本是需要的，而机会方法并不这样做。譬如，定额抽样保证样本中男士的百分比，等于总体中男士人口的百分比。而在概率抽样下，我们只能说样本中男士的百分比很可能接近于总体中的百分比：使必然性变为可能性。但判断加选择往往带偏性，而机会是不带偏袒性的。这就是为什么概率方法比判断方法奏效的原因。

为极小化偏性，就应该使用不带偏袒性且客观的概率方法选取样本。

表 4　Gallup 民意测验在 1948 年后的总统选举中的记录

年份	样本容量	获胜侯选人	Gallup 民意测验预测值	选举结果	误差
1952	5 385	艾森豪威尔	51%	55.4%	+4.4%
1956	8 144	艾森豪威尔	59.5%	57.8%	-1.7%
1960	8 015	肯尼迪	51%	50.1%	1%的+0.9
1964	6 625	约翰逊	64%	61.3%	-2.7%
1968	4 414	尼克松	43%	43.5%	1%的+0.5
1972	3 689	尼克松	62%	61.8%	-0.2%
1976	3 439	卡特	49.5%	51.1%	+1.6%
1980	3 500	里根	55.3%	51.6%	-3.7%
1984	3 456	里根	59.0%	59.2%	1%的 0.2
1988	4 089	布什	56.0%	53.9%	-2.1%

注：百分数是主要政党得票数的；在 1968，Wallace（华莱士）用主要政党计算：误差是"预测一实际"民主党得票数的差。来源：Gallup 民意测验（美国公众意见研究所）。

6. 关于盖洛普民意测验的仔细剖析

即使选取样本时用的是概率方法，某种程度的偏性几乎是不可避免的，这给调查机构带来许多实际困难。这里将围绕 1984 年选举中 Gallup 民意测验所使用的问卷组织讨论。见图 3，4 和 5。

不投票者。在美国的典型的总统选举中，约有三分之一到一半的合法选民没有参加投票。Gallup 民意测验是猜测选民的选举意向，因此不投票者与他们无关，应尽可能排除在样本之外。这并不容易办到，由于不投票与某种耻辱感相联系，因此，许多回答者即使明知自己不会去投票但仍说他们将会去投票。筛选不投票者这一疑难问题由问卷的题 1—6 以及其后的某些问题处理：这是民意调查者的热门话题。譬如，问卷题 3，问回答者将到哪里投票（图 4）；知道答案的回答者很可能会去投票。题 13 问回答者上届选举投票了没有，这样措词是为了使否定性回答易于作出（补偿不投票所带的耻辱感）。上届选举参加投票的人，这次很可能也参加投票



“我说，我大概百分之四十二赞成 Nixon（尼克松），百分之三十九赞成 Rockefeller（洛克菲勒），还有百分之十九尚未决定。”

（见图 5）。

这组问题是用来决定回答者是否可能参加投票的，选举结果的最终预测只能基于那部分被判断很可能参加投票的样本。谁在每次选举中确实参加了投票是有记录可查的。Gallup 机构对选举以后的研究表明预测哪些人将参加投票是相当准确的。其研究也同时证实筛选掉很可能不参加投票的人提高了选举预测的准确率，因为很可能参加投票的人的意向与很可能不参加投票人的意向有很大差别[15]。由于 Gallup 民意测验筛选了绝大部分不投票者，样本中留下的少数几个不会对估计有太大影响；因此他们的存在只导致结果很小的偏性。

未决定者。被访问对象中的某些百分比的人尚未决定将如何投票。问卷题 7，是询问有关投票意向的，其构思是为了使上述百分比保持一个尽可能小的值。首先，它问回答者在访问的当天，而

不是选举日，会投谁的票？对尚不能做出决定的回答者，请他们对“今天你更倾向于哪一位侯选人”表示意见。最后工具是印有侯选人名字的无记名选票(图3)。回答者不需要开口说出他们的意向，只需在选票上打上记号并投入访问人员随身携带的盒子中。

图 3 Gallup **民意测验无记名选票**，1984。

访问人员用无记名选票使未决定的回答者人数极小化。

民主党 □蒙代尔 与弗拉罗	共和党 □ 里根 与布什

图 4 Gallup 民意测验 1984 年选举问卷

调查：A1813

日期：1984. 10. 25



美国与世界

Gallup 调 查

Gallup 机构有限股份公司对本问卷及其书面或口头形式的回答，从未授权出版发行，复印，传播或其它用途。违者将追究其全部法律责任

由主要报社、公司和机构资助 Gallup 有限股份公司，普林斯顿，新泽西州 08540，1977 版权

自我介绍：

我正在进行 Gallup 调查，我乐于听取你关于在某些感兴趣的话题上的观点。

开始时间：____

结束时间：____

历　　时：____

1. 你对即将到来的十一月份的选举有多关心——很多或仅一点？

　1. □很多　　2. □有点——(自动提出)，

　3. □很少　　y. □不

2. 你在这个选举分区或选区投过票吗？

　1□是，　　2. □不，　　y. □不知道

3. 住在这附近的人到哪里投票？

　1. □详述：________________

y. □不知道

4a. 你现在是否已登记,从而能在今年 11 月份的选举中投票?

1. □是——(转到问卷题 5), 2. □不

3. □不必登记——(转到问卷题 5), y. □不知道。

4b. 你是否打算去登记,从而能在今年 11 月份的选举中投票?

1. □是, 2. □不, 3. □其它:______________

5. 一般地说,你对政治有多感兴趣—— 大量,相当,仅略有些或一点也不感兴趣?

1. □大量 2. □相当, 3. 略微, y. 不。

6. 你有多少时间谈论你投票的事—— 总是,几乎总是,有时或很少?

1. □总是, 2. □几乎总是, 3. □有时,

4. □很少, 5. □其它:__________________

y. □从不

7. 假设今天你投票选举美国的总统和副总统。这是一张印有这些职务侯选人名字的 Gallup 民意测验无记名选票。(撕下选票并交给回答者。)烦请在选票上标出你今天所赞同的侯选人,并将折叠好的选票投入盒中。

访问员:若回答者将选票退回并表示他尚未拿定注意或拒绝在其上打标记。可说:啊!你是否愿意在选票上标出你今天倾向哪一位侯选人?若回答者仍不能决定或拒绝在选票上做标记,请将情况在选票上写明并务必将选票投入盒中。

8. 此刻你对自己的选择有多坚定—— 非常坚定,相当坚定或毫不坚定?

1. □非常坚定, 2. □相当坚定,

3. □毫不坚定, 4. □不选择, y. □不知道。

9a. 你本人打算在今年 11 月份的选举中参加投票吗?

1. □是

2. □不 ┐
y. □不知道 ┘(转到问卷题 10a)

9b. 你有多么肯定你将参加投票——绝对肯定,相当肯定,或不肯定?

1. □绝对, 2. □相当, y. □不肯定

10a. 若今天举行国会选举,你乐于见到哪一政党在本众议员选区获胜,民主党或共和党?

1. □民主党
2. □共和党 （转问卷题 11）
3. □其它

y. □未决定或拒绝回答。

10b. 就今天来说，你更倾向于民主党或共和党？

1. □民主党，　2. □共和党，　3. □其它

4. □未决定，　y. □拒绝回答

11. 这是一张阶梯图(将卡片 1 交给回答者。)若设标有 10 的阶梯的顶端(点)表示明确肯定今年 11 月份选举中将参加投票的人，而标有 0 的阶梯的低端(点)表示明确肯定将不参加投票的人。你认为自己应安放在阶梯的哪一位置上？(访问员：在数上打圈)

10　9　8　7　6　5　4　3　2　1　0

y. □不知道

这些方法是用来极小化未决定者的百分比的。但仍难免有漏网的，若认为他们很可能参加投票，则 Gallup 民意测验就必须推测他们会怎么投。利用问卷题 12—14(见图 5)可收集到有关政治态度的某些信息。这些信息可用来预测未决定者将如何投票，当然还很难说预测奏效有多好。

回答偏性。问卷题的措词，甚至访问员的声调或姿势都会在某种程度上影响被访人做的回答。这类失真称为回答偏性。1984 年的选举调查中就有一个令人吃惊的例子：改变侯选人名字的顺序后有约 5%的回答出现变化，一般名字排在第一位的侯选人较为有利。为了控制回答偏性，所有访问者要用相同的问卷，且访问程序也要尽可能标准化。无记名选票方法有利于减少访问员的政治态度对被访问对象回答的影响。

不回答偏性。即使亲自访问，仍会遗漏许多对象。由于这些人趋于与能访问到的对象有所不同，从而产生了不回答偏性。在某种程度上，这种偏性可以校正，只要对那些可访问到但难于得到其回答的对象加较大的权则可。问卷题 20 可用来获取这方面的信息。它是询问对方前些日子是否在家的。这做得很巧妙，正象你们阅读

问卷时可获悉的那样。

家庭偏性。Gallup 民意测验只访问被选出供调查家庭的一个成员。这种做法歧视生活在大家庭中的人，他们在样本中没有足够的代表性，因此应给被选为代表的那些人加较大的权。家庭的大小由问卷题 18 获得。

检验数据。在 1984 年，年满或超过 25 岁的美国公民中约 75%完成高中或大学学业。Gallup 样本中往往包含过多这类受过良好教育，且倾向于共和党的人。在详细分析中，较大的权加给那些受教育较少的访问对象的回答(问卷题 16)上。其它人口统计的数据可类似地使用。这种方法称为“比估计”。切莫混淆比估计(一种好的方法)与定额抽样(一种差的方法)。比估计是应用于样本已被抽取之后的一种客观的算术方法以补偿抽样过程中各种小的偏差。定额抽样是一种选取样本的方法。它有较大的主观因素——在访问员选择他访问对象时——且导致较大的偏性。

对访问员的控制。在大规模的调查工作中，总存在审查访问员是否遵循指示的问题。因此在问卷中设有一些额外问题，从而可检查其回答来审核一致性；不一致表示访问员没做好工作。作为工作质量的进一步考核，管理人员会再次访问被访问者中的一小部分人。

交谈是贬了值的。根据人们告诉访问员他们打算做什么来预测他们在选举日将做什么是要冒点风险的。人们可能并不愿意泄露他们的真正意向，即使愿意，过后他们仍可能改变主意。说的与做的有时并不一致这是经验事实。

图 5 Gallup 民意测验 1984 年选举的问卷，(续)

好！下面有几道问卷题，目的是使我的公司能留意与我交谈的人们的剖面。

12. 到今天为止，你认为自己在政治上是一名共和党人，民主党人或无党派人士？

1. □共和党人，　2. □民主党人

3. □无党派人士，　4. 其它：________

13. 在 1980 年 11 月份的选举中——当 Carter（卡特）撞上 Reagan（里根）和 Anderson（安德森）时——是否有什么事发生使你没能参加投票，或你去投了票？投了谁的票？

1. □Carter　2. □Reagan　3. □Anderson

4. □其它　5. □投了票，记不起来投了谁的票，

6. □不，没投票，　y. □记不起来是否投过票。

14. 你或你的（丈夫/妻子）是工会会员吗？

1. □是，回答者是，

2. □是，配偶是，

3. □是，都是，

y. □不，都不是。

15. （将卡片 2 递给回答人）请告诉我这卡片上那一类目最贴近地描述了你们家庭中主要雇佣劳动者的工作类型，只须报出号码即可。（访问员：若主要雇佣劳动者失业，则询问若被雇佣，他/她将从事哪一类工作。）

1. □，　2. □，　3. □，　4. □，　5. □，　6. □，　7. □，

8. □，　9. □，　10. □，　11. □，　12. □，　13. □，

14. □，　15. □，　16. □其它：________，　17. □说不上。

16. 你在学校读完的最后学历或最后年级是什么？

1. □无或 1－4 年级，　2. □5，6，7 年级，

3. □8 年级，　4. □高中没毕业（9－11 年级），

5. □高中毕业（12 年级），　6. □技术，工艺或商业专科，

7. □学院，大学没念完，　8. 学院，大学毕业。

17. 你的宗教信仰是什么——基督教，罗马天主教，犹太教或东正教诸如希腊或俄罗斯东正教？

1. □基督教，　2. □罗马天主教，　3. □犹太教，

4. □东正教　5. □其它：________　y. □无。

18. 目前生活在这家庭中年满或超过 18 岁的成员包括你本人有几位？包括房客，佣人或其它生活在这家庭中的雇工，（在数上打圈）

1，2，3，4，5，6，7，8，9 或更多

19. （递给回答人卡片 3）你主要是哪个或哪几个民族的后裔？只需报号码则可。

1.□，　2.□，　3.□，　4.□，　5.□，　6.□，　7.□，　8.□，　9.□，

10.□，　11.□，　12.□，　13.□，　14.□不知道(自愿表示)或拒绝。

20a. 我们对人们是否常在家收看电视或听收音机很感兴趣，你能告诉我昨天(昨晚/上星期六)这个时侯你是否在家？

(访问员：看着访问员手册提这问题。)

1.□是，在家，　2.□不，不在家。

20b. 前天(晚上/星期六)这个时侯你也在家吗？

1.□是，在家，　2.□不，不在家

20c. 前天(晚上/星期六)这时侯的情况如何？情况是____

1.□是，在家

2.□不，不在家

21. 你的年纪多大？

记下年龄：________

22. 检查一下是否

1.□白人男性

2.□白人女性

3.□黑人男性

4.□黑人女性

5.□其它种族男性(详细说明)__________

6.□其它种族女性(详细说明)__________

为了便于我的公司在它需要的时侯能检查我在这次访问中的工作，是否可请你留下你的姓名，地址和电话号码？

姓名：________

地址：____________________

房子号码，路名，房间号码

城市：________州：________邮编：________

电话：分区________电话号________y.□无电话

在此填上访问员的代号	特此宣誓这是一次真实可靠的访问 __________(访问员签名) 访问日期：__________　访问结束时间：__________

7. 电话调查

现在许多调查是用电话进行;费用上的节约是惹人注目的,且——若做得对——结果是好的。1988 年,Gallup 民意测验用电话实施它的预选调查,从 Nebraska(内布拉斯加)州 Lincoln(林肯)市的一个房间里,200 名访问员在三天内采访了全国。

他们是怎样选取样本的呢? 调查工作人员把电话号码看作有一个电话局(地区号加头三个数字)和一个号码库(紧挨其后的二个数字);然后再外加二个数字构成号码。在电话调查中,第一阶抽取一个电话局样本;第二阶从已选出的电话局抽取库样本[16]。最后阶段随机抽取二个数字构成电话号。这种方法常称为随机拨号,或简称为 RDD。

电话局	库	数字
415—767—	26	76

没装电话的人必与我们其余的人有区别,且导致电话调查中的偏性。由于现在几乎家家都装了电话,因此它的影响是小的。另一方面,约有三分之一住户的电话未入册。富人和穷人的电话偏向于未入册,故在电话簿里中产阶层过多载入。根据电话号码簿来抽样将会产生真正的偏性;而随机拨号回避开了这一问题。

在实际中,某些号码库几乎为全商行使用。电话公司保留许多其它的库供将来使用。因此,Gallup 民意测验只取有 3 个以上住户电话号的库。若你随机取一个电话号,它当时恰好分给住户的机会低于 20%;用 Gallup 民意测验所使用的方法,机会上升为 80%。某些商行的电话号仍将会在样本中出现,因此计算机将检查黄色的专页(译者注:指电话号码簿中商行号码部分),并将列在它上面的号码删去:Gallup 民意测验是抽人不是抽商行。

象往常一样不回答产生问题。因此,Gallup 民意测验的大部分访问都在傍晚和周末,人们更可能在家的时侯进行。若没人接电话,访问员还会再次拨打至三次之多。(有的反复拨打高达 15 次.

那当然更好，但费用比较贵。)Gallup 处理的不回答率约为 35%，这可与亲自访问率相媲美，费用却只为其三分之一左右，预测达到的指标也相当好(表 4)。

这就是调查机构使用电话的原因。

8. 机会误差

前几节指出了实际调查机构所面对的现实困难：人们不在家，或即使他们在家但不愿暴露他们的真实意向，或他们虽然做了但尔后又改变了主意。如果没有这些困难，情况会是怎样呢？设想有一只装有大量票子的盒子，这些票子有的标 1，有的标 0，这是总体。某调查机构受雇估计盒中 1 的百分数，这是参数。该机构从盒中随机不放回地抽取 1 000 张票，这是样本。这里不存在有关回答的问题：票子全在盒中。随机抽取消除了选择偏性。票子自己不会在 0 和 1 之间变来变去。从而，样本中 1 的百分数将是盒中 1 的百分数的一个好的估计。但仍会有少许偏离，原因是样本仅为总体的一部分。但由于样本是随机挑选的，偏离量由机会控制：

样本中 1 的百分数＝盒中 1 的百分数＋机会误差

在更复杂的情况下，等式中必须考虑偏差：

估计＝参数＋偏差＋机会误差

现有若干有关机会误差的问题：

- 它们可能会有多大？
- 它们对样本大小的依赖程度为何？对总体大小呢？
- 为了将机会误差置于控制之下样本容量必须取多大？

这些问题将在下面二章中回答。

习题 A

1. 民意测验经常用电话实施预选调查。这会使结果有偏差吗？怎样造成的？若样本是从电话簿中抽取时情况又将如何？

2. 大约在 1930 年，在纽约进行了一项关于前黑人奴隶对他们主人及奴隶状态条件的态度的调查[17]。访问员中有些是黑人，有些是白人。你认为这二组访问员会获得类似结果吗？说出你的理由。

3. 关于奴隶制的一项研究估计“11.9%的奴隶是熟练手艺人”。这估计是根据 Louisiana（路易斯安娜）州 Plaquemines Parish 的 30 个农场的记录作出的[18]，它可信吗？

4. 1968 年在荷兰进行了一项研究，将 18 岁男孩的智力与他们兄弟姐妹的人数联系起来[19]。在荷兰，所有 18 岁的男人都要接受兵役预征检查。这项检查包括以“Raven 的进展基质试验”著称的智力测验，以及诸如家庭大小等有关人口的统计变量的问题。研究将智力测验得分与家庭的大小联系起来。研究使用了 1968 年的全部检查记录。

 (a)什么是总体？什么是样本？

 (b)存在机会误差吗？

5. 在一项研究中，教育检测行政机构需要一个有代表性的大学生样本[20]。为了抽取样本，他们首先将全体学院和大学划分成相对同类的组（一个组由学生人数不少于 25 000 人的公立大学组成，另一组由学生人数不多于 1 000 人的私立四年制学院组成，等等）然后，他们根据自己的判断从每一个组中选出一所有代表性的学校。产生出学校的样本。接着再要求样本中的每所学校抽取学生样本。这是一种获得有代表性的学生样本的好方法吗？

6. Gallup 每月一次的民意测验是基于“经科学选取作为美国公众有代表性的剖面”的约 1 500 人组成的样本。Gallup 民意测验认为他们的样本是有代表性的主要是因为，

 (i)样本在诸如种族、性别、年龄、收入以及受教育程度等这些特征方面与总体相似，

 还是

 (ii)样本是用概率方法选取的。

7. 在某所大学进行一项调查以估计本学期住在家里的大学生的百分数。总体是什么？参数是什么？

8. 注册管理员手上有一份全体大学生按字母顺序排列的名单与现住址。设本学期有 10 000 名学生，某人建议从 1 到 100 中随机取一个数，然后在名单上从这个数往下数下去，取这名字及其后每第 100 位的名字组成样本。

(a)这是一种概率方法吗?

(b)它与简单随机抽样相同吗?

9. 40 年代末,对 Welsh(威尔斯)一个开采煤矿的镇中胸部疾病盛行进行了研究;600 名自愿者接受了胸部 x—射线检查[21]。当时的两种主要胸部疾病是矽尘病(由于尘埃的吸入造成的肺部组织疤痕斑斑)和肺结核。数据按自愿者呈递的先后次序进行分析处理。前 200 名对象中经检查患有肺结核的百分数与最后 200 名的百分数____________________。请在空格处填入下面二组短语之一:

几乎相同　　　　颇为不同

解释你的理由。

10. 电视广告拍卖受 Nielsen(尼尔逊)收视率的强烈影响。在它的年度报告中,Nielsen 机构并不述说他们是怎样选取样本的。报告上说[22]:

> Nielsen,今天和以往一样,奉献于使用最新,最可靠且经彻底考验过的研究技术。这是我们对我们通过电视、有线电视和广告界为之服务的人们的一种承诺。…
>
> 本册子中的 Nielsen 数据是电视观众和电视使用的其它特征的估计,是用 Nielsen 电视索引和 Nielsen 电台索引中的度量导出。这里数学措词的使用请务必不要当成 Nielsen 表示这些度量是精确的数学值…

简短地评论之。

11. 1988 年 9 月 11 日,星期日,The San Francisco Examiner(《旧金山检查者》)刊登了一则传奇消息,标题是

10 名生物教师中有 3 人支持《圣经》的创世说。

> Arlington,Texas(阿灵顿,德克萨斯)　根据星期六公布的一项全国性调查,接受民意测验的百分之三十的中学生物教师相信《圣经》的创世说,又百分之十九错误地认为人类与恐龙曾生活在同一年代。
>
> "我们在生物学教学中做了某些非常、非常、非常错误的事",阿灵顿,德克萨斯大学两位社会学家之一的 Dana Dunn(德纳·德昂)说,Dunn 和 Raymond Eve(雷家·伊夫)将问卷寄给了 20 000 名高中生物教师,他们是从全国科学教师协会提供的名单中随机选取的,并收到 200 份答卷。…

两位社会学家在统计工作上做了某些非常、非常、非常错误的事,那是些

什么?

这些问题的答案见708—709页。

9. 复习题

复习题可能包含前面几章的内容

1. 为了度量广告宣传对某一种商标洗衣粉的效应,做了两种调查[23]。在第一种调查中,访问员询问家庭主妇是否使用这种商标的洗衣粉。在第二种调查中,访问员要求看一下使用的是哪一种洗衣粉?你认为这两种调查会导致类似结论吗?说出你的理由。
2. 关于奴隶制的一项研究估计一个奴隶每年只有2%的机会被卖转入州际贸易。这估计是根据马里兰州Anne Arundel县的拍卖记录做出的[24]。它可信吗?说出理由。
3. 在一项研究中,需要抽取一个住在旧金山的日裔美国公民的有代表性的样本[25]。抽样程序如下:经与日裔社团中有代表性人物磋商后,选出该市日裔地区的四个最具代表性的街区;并将居住在这四个街区的全体居民取为样本。然而,与人口普查数据相比较后表明该样本没有包含大学程度日裔的足够比率。这该如何解释?
4. (假设的。)规划部门进行一项调查以确定某市家庭大小的分布。他们抽取了一个1 000个家庭的简单随机样本;但是经过多次拜访后,访问员发现样本中仅853家有人在家。规划人员抽取第二批家庭以免面对如此高的不回答率,并将第二批中完成访问的前147家凑足样本达到1 000个家庭的规划强度。他们数得这1 000个家庭中共有3 087人,并估计该市家庭平均大小约为3.1人。这个估计可能是太低、太高或基本是对的?请解释。
5. "Ecstasy"在80年代中期是一种流行的迷幻毒品。它产生一种称为"yuppie high"的可笑的欣快感。一名调查人员进行一次仔

细的调查以便估计斯坦福大学服用这种迷幻毒品的流行情况。他派了两名助手到主要校园广场并指示他们询问在特定时刻路过的全部学生。结果被询问的369名学生中39%说他们至少服用过一次Ecstasy[26]，调查人员的程序给出了斯坦福学生的一个概率样本吗？回答是或不是，并解释之。

6. 一币硬币抛1 000次。有两种选择：

(i)若出现头像的次数在490次到510次之间时赢1美元；

(ii)若出现头像的百分数在48%到52%之间时赢1美元。哪一种选择较好？或它们相同吗？解释之。

7. 设将一枚硬币抛100次。你可以选11个数。若头像出现的次数等于你选出的11个数中的一个，你赢一美元。你会选哪11个数，你赢的机会(近似地)有多大？解释之。

8. 一魔术师将一块M&M(一种巧克力——译者注)藏入有无限多只的一排盒子中的一只里。

……	□	□	□	□	□	……
	−2	−1	0	+1	+2	

你想尽可能快地找出它。你实际上只有时间查看11只盒子。该魔术师说他会给你一点提示，在一瞬间，他很快地抛一枚硬币100次并数出头像出现的次数。但他不告诉你这个数字，也不告诉你放有M&M的盒子的号码数，而只告诉你这二个数的和。

(a)若这个和是65，你会查看哪11只盒子？

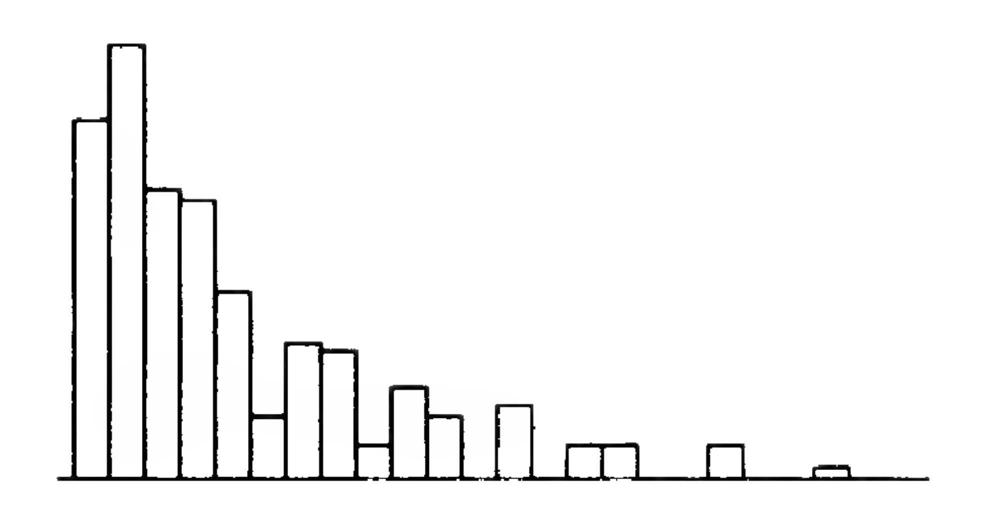
(b)问题如(a)，仅将65改为95。

(c)一般规律是什么？

(d)根据这一规律，你可能会怎样放置M&M？

9. 下面这张图是一张概率直方图吗？或是一张数据的直方图？或根据给定信息能确定这张图吗？

10. 一家医院一月份有218名婴儿诞生[27]。另一家有536名。哪一家更可能有55%或以上的男婴？或者它们是等可能的？解释之(一个初生婴儿为男婴的机会约为52%。)



10. 小结

1. 样本是总体的一部分。

2. 参数是总体的数字特征。一般参数不能精确确定，而只可能估计。

3. 统计量可由样本计算，并可用来估计参数。统计量是调研人员知道的；参数是调研人员要知道的。

4. 当估计参数时，一个主要的问题是精确度：估计量将会多接近？

5. 某些选样本的方法可能产生精确的估计。而其它方法被选择偏性和不回答偏性破坏。因此，当考虑抽样调查时，问一下自己：

- 总体是什么？参数是什么？
- 样本是如何选取的？
- 回答率是多少？

6. 大样本并不能防止偏倚

7. 在定额抽样中，样本由访问员精心选出以便在某些关键方面与总体相似。这种方法似乎合乎逻辑，但常给出坏的结果。理由是访问员方面的无意的偏性。

8. 抽样的概率方法用一种客观的机会过程选取样本，而不给访问员留有任何自主处理的权力。概率方法的标志：调研人员可以算出总体中任一个体被选入样本的机会。

9. 一种概率方法是简单随机抽样，是指随机不放回地抽取对

象。

10. 利用概率方法获得的样本很可能具有代表性：没有主观意识的机会是不带偏性的。

11. 即使用的是概率方法，由于偏差和机会误差，估计可能有点偏离参数：

估计＝参数＋偏差＋机会误差。

20
抽样中的机会误差

谨致全体在座的和某些没有到的女士们

——Jerzy Neyman 常提议的祝酒词

1. 引言

抽样调查含有机会误差。本章将讲解对于成分为已知的总体中进行简单随机抽样，如何求出机会误差可能占的百分比，这主要依赖于样本的容量，而不是总体的大小。先看个例子。

一项健康研究是基于 6 672 名 18 到 79 岁之间的美国人组成的有代表性剖面上的。现在一位社会学家希望对这些人作跟踪研究。她没有足够财力对他们全体进行研究；事实上，她只有研究他们中一个 100 人样本的经费。为了避免偏差，她打算随机抽取样本。在如下进行的虚构对白中，她正与她的统计员讨论这个问题[①]。

社会学家（下面简记为社）　我想我必须把全部 6 672 个名字分开写在票子上，将它们放进盒子中，并随机抽出 100 张。听起来象有大量工作。

统计学家（下面简记为统）　我们在计算机里有编号由 1 至

6 672 的号码。因此，你只要从这范围内随机抽取 100 个数，你的样本就是具有这些编号的人。

社. 是的，但是我还是不得不将数 1 至 6 672 写在票子上，你没省我多少时间。

统. 那并不是我所想的。事实上，这么大的一只盒子，你很难把票子混匀。如果你不能混匀，你抽出的绝大部分票子可能是你在最后放进盒中的。那将是真正的偏差。

社. 那么，你建议干什么？

统. 计算机有随机数发生器。它会帮我们从 1 到 6 672 中随机取一个数，具有该编号的人进入样本。然后，它将随机取不同于第一个数的第二个编号。那就是第二个人进入样本。它照这样不断地取下去直到取足 100 人为止。你不必去混匀票子，由随机数帮你混匀。另外，它们还免去了全部书写工作。

社. 好。但是，如果我们用计算机，你能保证我的样本有代表性吗？

统. 你指的是什么？

社. 噢，譬如在原来的研究中，有 3 091 个男的和 3 581 个女的：46%是男的。我希望我的样本中有 46%的男性。除此之外，我希望他们的年龄分布也合适。然后，还要考虑收入和受教育的情况。当然，我真正要的是一个组，他们对保健的态度具有代表性。

表 1 随机选取 100 人，并按性别分类。51 人是男的(M)，49 人是女的(F)。在总体中，百分数是 46%和 54%。

F F M F M	M F M M M	M F M M M	M F M F F
F M M F M	F F M F F	M M F F F	M F M F M
F M T F M	M F M M F	M F M F M	M M F F F
F M M M F	M F M M F	M M M M F	F F M F M
F M F M M	M F F F F	M M F M M	F F F F F

统. 眼下先别操之过急。首要的先来，我抽组样本给你看。看表1，计算机选出的第一个人是女的，第二个也是，但第三个是男的，等等。总起来，你得到 51 个男的。相当接近。

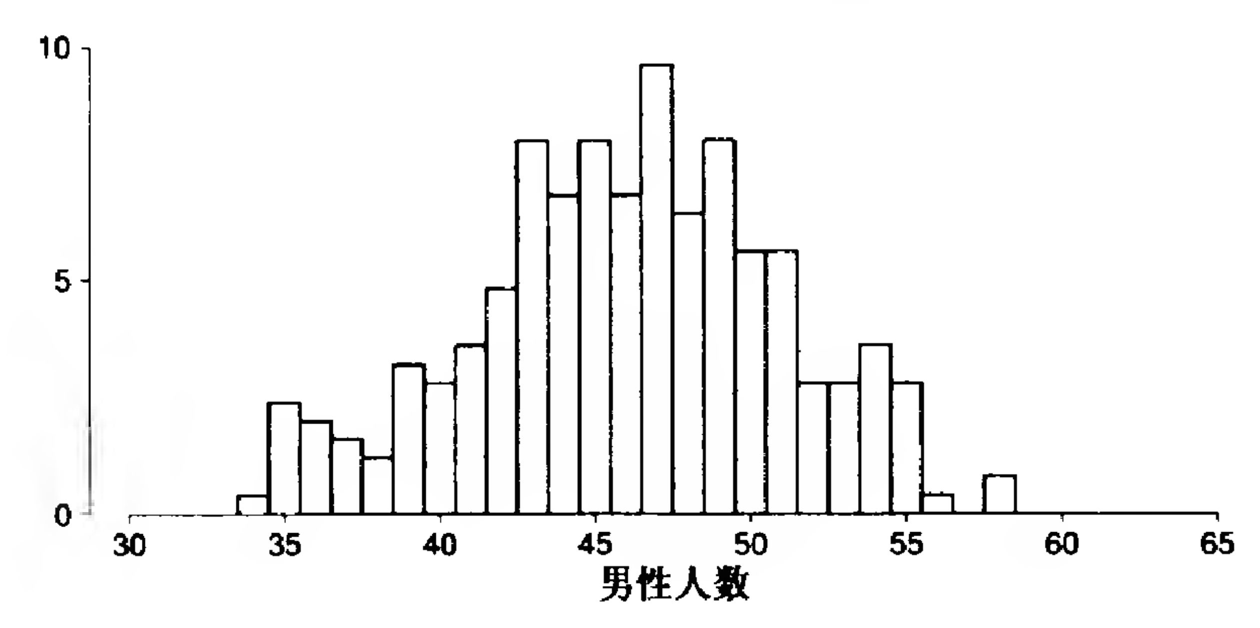
社. 但应该只有 46 个男的，这样本的比率不对。计算机一定出了什么毛病。

统. 不，事实并非如此。别忘了，样本中的人是随机抽取的。仅靠抽取时的运气，你取到男的可能过多——或过少。我让计算机给你取一大批样本，总共 250 个（见表 2）。男的人数变化范围从最低 34 到最高 58。这一大批样本中仅 17 组样本恰好有 46 个男的。这里有张直方图（见图 1）。

表 2 从某项健康研究回答人（其中 46% 为男的）中抽取的二百五十个样本（每个样本 100 人）。各样本中男性人数如下。

51 40 49 34 36	43 42 45 48 47	51 47 50 54 39	42 47 43 46 46	51 43 53 43 51
42 49 46 44 55	36 49 44 43 45	42 42 45 43 55	53 49 46 45 42	48 44 43 41 44
47 54 54 39 39	52 43 36 39 43	43 46 47 44 55	50 53 55 45 43	47 40 47 40 51
43 56 40 40 49	47 45 49 41 43	45 54 49 50 44	46 48 52 45 47	50 53 46 44 47
47 46 54 42 44	47 47 36 52 50	51 48 46 45 54	48 46 41 49 37	49 45 50 43 54
39 55 38 49 44	43 47 51 46 51	49 42 50 48 52	54 47 51 49 44	37 43 41 48 39
50 41 48 47 50	48 46 37 41 55	43 48 44 40 50	58 47 47 48 45	52 35 45 41 35
38 44 50 44 35	48 49 35 41 37	46 49 42 53 47	48 36 51 45 43	52 46 49 51 44
51 51 39 45 44	40 50 50 46 50	49 47 45 49 39	44 48 42 47 38	53 47 48 51 49
45 42 46 49 45	45 42 45 53 54	47 43 41 49 48	35 55 58 35 47	52 43 45 44 46

图 1 容量为 100 的样本中男性人数的直方图



社. 什么东西在阻止男性人数出现46人?

统. 机会变异,记得前几天我跟你讲过的Kerrich试验吗?

社. 是的,但那是关于抛硬币的,不是抽样。

统. 噢,抛硬币与抽样之间没多大区别。每次你抛一枚硬币时,你要末得到头像要末得到背面。如果你计算头像数,次数要末增加1要末不变。每次的机会是50对50。它与抽样一样,每次计算机选出一个人进入样本,选出的要末是男的要末是女的,如果对男的计数,则个数要末增加1要末不变。每次的机会差不多是46对54——从盒中取100张票子不会很大程度上改变盒中的比率。

社. 事情的实质是什么?

统. 抽样中的机会变异恰如抛硬币中的机会变异。

社. 唔。若增大样本的容量会出现什么呢?样本会变得更象总体吗?

统. 对。譬如,假设将样本容量增大为4倍,或400。我让计算机抽另外250个样本,这次每组样本中有400人。这些样本中的某些男性的百分数低于46%,其它的高于46%。低的为39%,高的为54%。图2是一张直方图。你可以将它与容量为100时的直方图相比较,将样本容量增大为4倍后,在百分数中的机会误差的可能大小下降为原来的一半。

社. 你能把这机会误差变得更具体点吗?

统. 让我写个等式:

样本中的百分数=总体中的百分数+机会误差。

当然,不同的样本,机会误差也不一样——记住表2中的变异性。

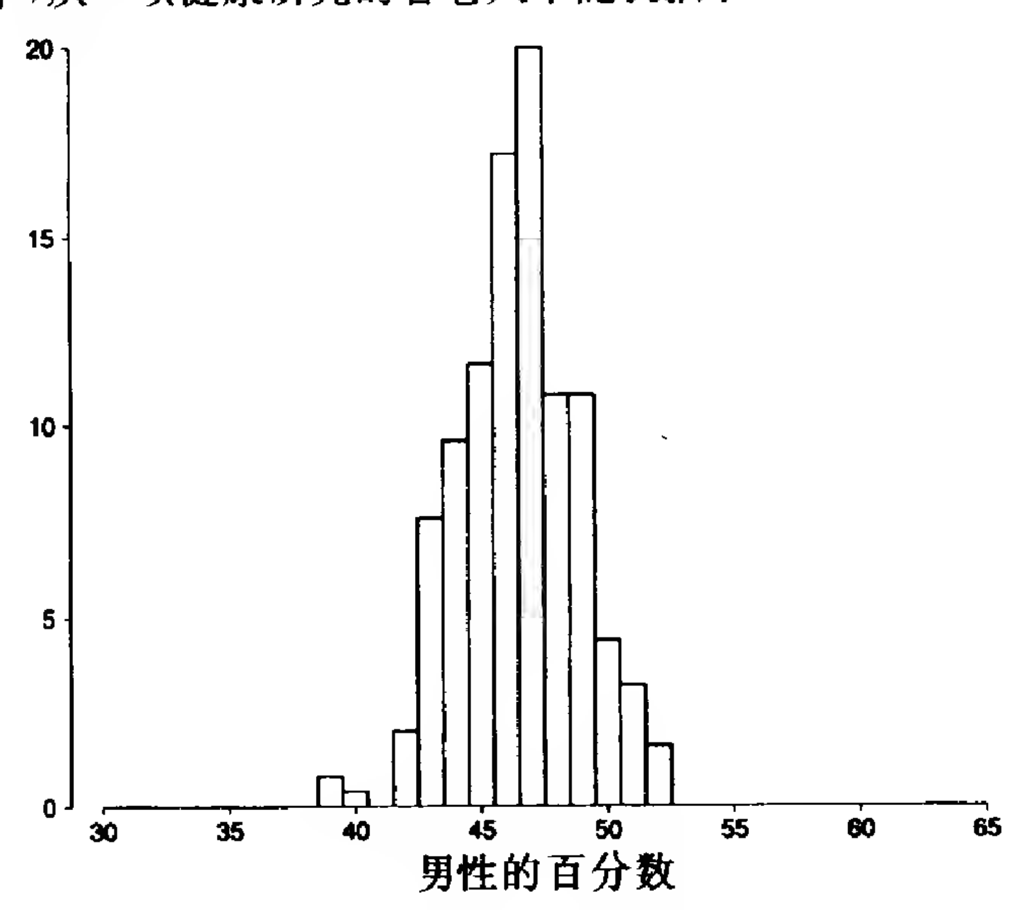
社. 好,如果我让你给我抽一个样本,用这个随机数思想,你能说出我的机会误差将有多大?

统. 不精确地,我能告诉你它可能有多大。如果你让我构造一个盒子模型,我可以算出标准误差,然后…

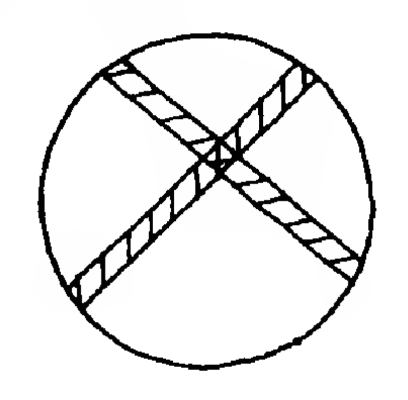
社.　等一等，有一点我先前给忘了，你怎么能有 250 个各有 100 人的样本？我的意思是，250×100＝25 000，而我们一开始只有 6 672 人。

统.　这些样本都是不相同的，但有某些人是他们共有的。看这张

图 2　在容量为 400 的样本中男性百分数的直方图。共有 250 个样本，从一项健康研究的答卷人中随机抽取。



草图。图的内部就象 6 672 人，有阴影的细条就象样本。细条是不同的，但它们相交迭。实际上，我们只对我们的抽样作了肤浅的探讨。容量为 100 的不同样本总个数超出 10^{200}。它等于 1 后面跟二百个 0。某些物理学家甚至不认为整个宇宙中会存在那么多的基本粒子。



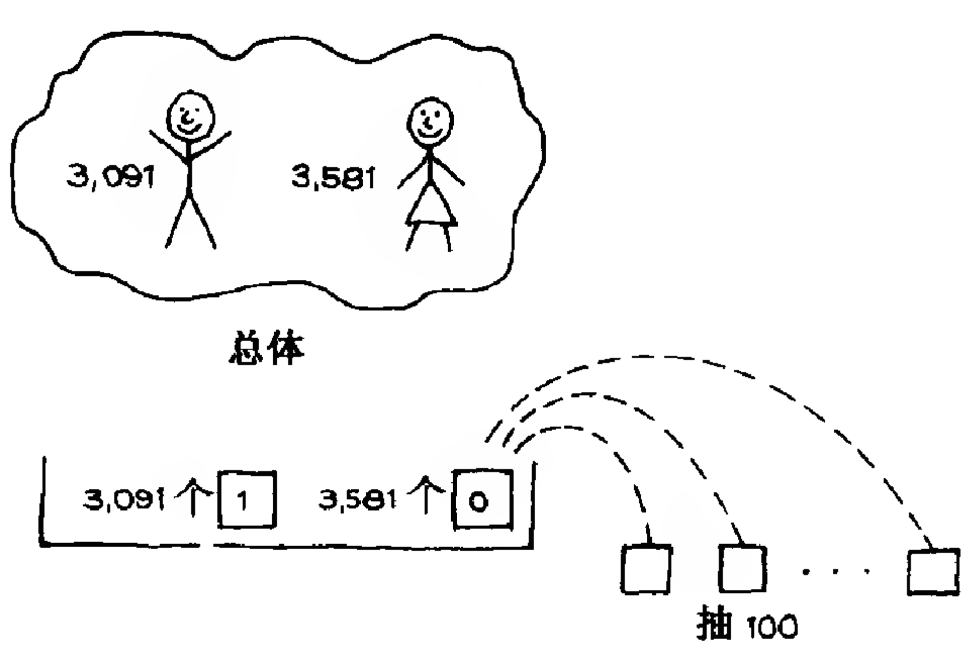
2. 标准误差

上节中的社会学家考虑在一项健康研究中从 6 672 名受试验者的总体中抽取一个容量为 100 人的样本。她认识到样本中男性的百分数不会与整个总体的百分数相等，会由于机会误差而偏离。这机会误差会有多大呢？答案由标准误差给出。对于社会学家的问题，标准误差是 5 个百分数点。换句话说，社会学家可以期望在她的样本中男性的百分数偏离总体中的百分数 5 个百分数点左右。现介绍计算这类标准误差的方法。想法是算出样本中男性的人数的标准误差 SE，然后再换算成相对于样本容量的百分比。样本的容量正是指样本中的总人数——在这里是 100。

为计算 SE，你需要一个盒子模型。例如，第 1 节中的社会学家从一个由 3 091 个男人和 3 581 个女人组成的总体中取一容量为 100 的样本。她将样本中的人按性别分类并对男人计数。因此，盒子中应该只装有 1 和 0（第 17 章，第 5 节）。样本中男性的人数等于从盒子中抽 100 张票所得数之和。

| 3 091 个 [1]　　3 581 个 [0] |

她采用简单随机样本，因此票子必然是不放回抽取的。这完成了盒子模型。



盒子模型

盒子里的 1 所占的分数是 0.46，因此盒子的 SD（标准差）为 $\sqrt{0.46\times0.54}\approx0.5$。100 张抽取之和的 SE 为 $\sqrt{100}\times0.5=5$。从盒子中 100 张抽取之和将在 46 附近，加或减 5 左右的误差。

换句话说，在社会学家的 100 人的样本中男性人数可望约为 46，加或减 5 左右的误差。男性人数的 SE 为 5。而 5 个在 100 个中占 5%。因此，样本中男人的百分数可望约为 46%，加或减 5%左右的误差，这 5%的加减误差量为样本中男人百分数的 SE。

为计算一个百分数的 SE，先求出对应数的 SE，再换算成相对于样本容量的百分比。

对较大样本会出现什么呢？例如，若社会学家取一容量为 400 的样本，样本中男性人数的 SE 变成 $\sqrt{400}\times0.5=10$。而 10 表示样本容量 400 的 2.5%，因此在容量为 400 的样本中男人的百分数的 SE 为 2.5%。样本容量乘 4，百分数的 SE 将除以 $\sqrt{4}=2$。

由于 SE 的公式中的平方根，计数的 SE 的增长比样本容量的增长要缓慢得多。因此，作为样本容量的百分数，计数的 SE 下降。大样本比小样本更为精确。

样本容量乘以某一因子，百分数的 SE 则除以这因子的平方根。（例如，样本容量乘以 4，百分数的 SE 则除以 $\sqrt{4}=2$。）这规律对放回抽样是精确的。即使对不放回抽样，只要抽取票子的张数与盒子中的票子张数相比要少得多，它也是一个好的近似。

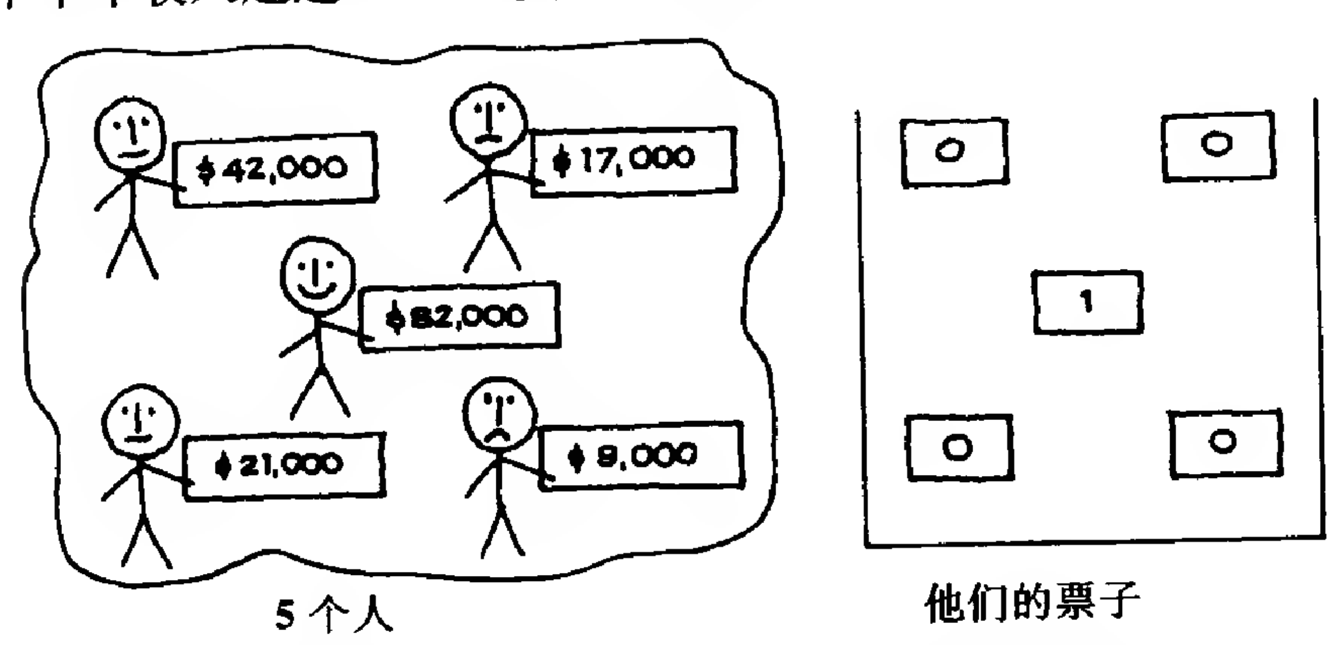
由于抽取的票子张数比起盒子中的票子张数要小得多，因此社会学家的 SE 的计算忽略了放回抽样和不放回抽样的差异。无论被抽出的是哪 100 张票子，盒子中 1 的百分数仍非常接近于 46%，就机会而言，放回和不放回抽样之间没有太大差异。

例 1. 在某镇，电话公司有 100 000 家用户，它计划从他们中抽取 400 户的简单随机样本，作为市场探索研究的一部分。根据人口普查数据，该公司的近 20%用户年收入超过 50 000 美元。求样

本中年收入超过这水平的人的百分数的 SE。

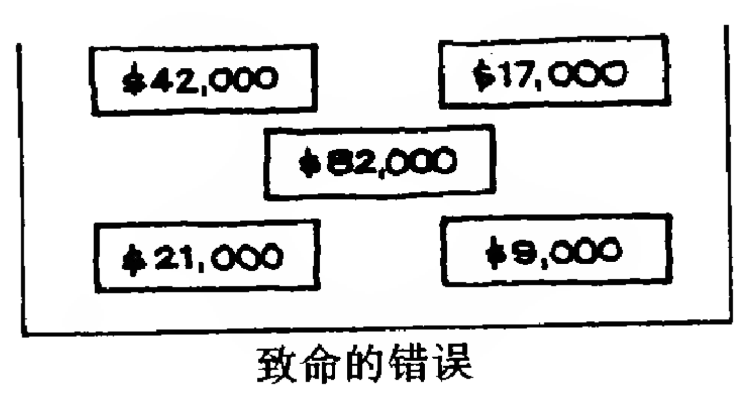
解. 第一步构造一个盒子模型。从总体中抽取 400 人的样本与从有 100 000 张票的盒子中抽取 400 张票子是相同的，总体中的每个人在盒子中都有一张票子与之对应，因此样本中的每一个人对应一次抽取。抽取是随机不放回的。

问题包括将样本中的人按年收入是否超过 50 000 美元分类，并对收入超过这水平的人计数。因此，盒子中的每张票子必须标上 1 或 0。收入超过 50 000 美元的人为 1，其余的为 0。已知总体中有 20%的人年收入超过 50 000 美元，故盒子中的 20 000 张票应标上 1，80 000 张标上 0。样本等价于从盒子中抽取 400 张票子。样本中年收入超过 50 000 美元的人数等于抽得数之和。



5 个人　　　　他们的票子

这完成了第一步，建立盒子模型。在这里，若票面上标的是每人的收入，则存在犯致命错误的可能性。模型就会回答样本中诸人收入总和这问题，而没有回答有关年收入超过 50 000 美元的人数的问题。



致命的错误

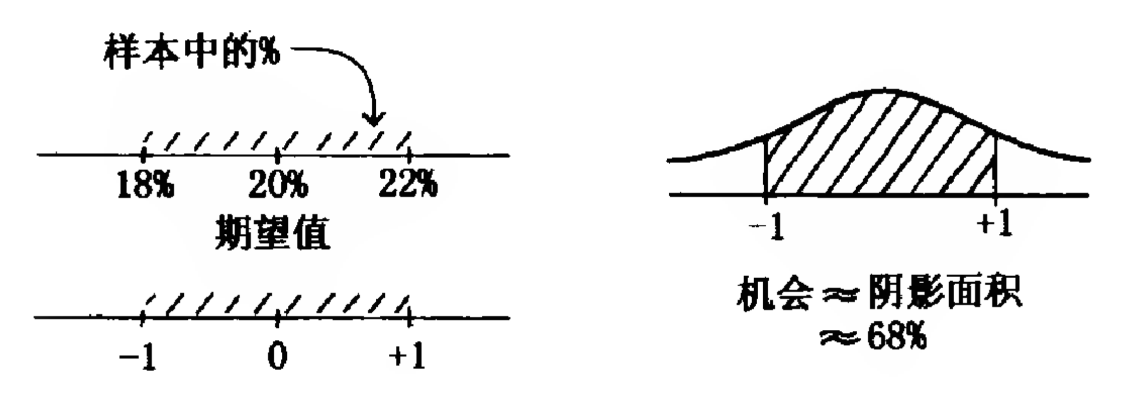
下一步是计算标准误差。首先需要盒子的 SD。它是 $\sqrt{0.2\times0.8}=0.4$，共抽了 400 次，故抽得数之和的 SE 为 $\sqrt{400}\times0.4=8$。抽得数之和将大约 $400\times0.20=80$，加减 8 左右的误差。换句话说，样本中年收入超过 50 000 美元的人数可望约为 80，加减 8 左右的误差。而 400 人（样本的容量）中的 8 人为 2%。因此，样本中年收入超过 50 000 美元的人的百分数可望约为 20%，加减 2 个百分数点左右的误差。这加或减的 2% 为百分数的 SE，完满结束本例。（可能不幸的是，统计学家用符号%同时表示“每一百个中有”和“百分数点”的缩写式。）

对充分大的样本，可以象第五部分中那样用正态近似：

- 机会误差以大约 68% 的概率小于 1SE；
- 机会误差以大约 95% 的概率小于 2SE；
- 机会误差以大约 99.7% 的概率小于 3SE。

例 2. 续上例，估计样本中有 18% 和 22% 的人年收入超过 50 000美元的机会。

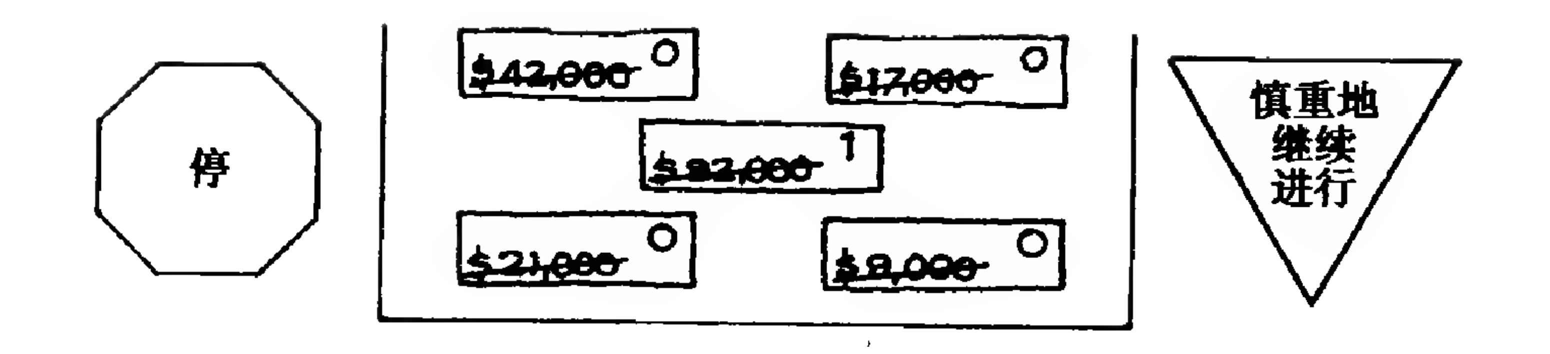
解. 样本百分数的期望值为 20%，SE 为 2%，现换算成标准单位：



解毕。

样本中人们的收入开始是定量数据——数。但是，例 1 和例 2 都包括分类和计数：对每个人都按年收入超过或不足 50 000 美元分类。然后再对高收入者计数。换句话说，数据被当成定性的：每一收入不是具有就是不具有年收入超过 50 000 美元这一性质。

什么时侯你可以换成 0－1 盒子模型？为回答这个问题，需要



注意对样本值所做的算术运算。计算可以包括：

- 将样本值加起来，求得一个平均数；

或

- 分类并计数，求得一百分率。

若问题包括了分类和计数，在盒子中放入 0 和 1（第 17 章，第 5 节）。

习题 A

1. 盒中装有 4 颗红弹子和 1 颗蓝弹子。从盒中随机放回地抽取 400 次。在抽出的样品中，红弹子的百分数约为____%，加减____%左右的误差。
2. 根据人口普查，某镇有 100 000 个年满 18 岁或以上的人。他们中 60%已婚，10%年收入超过 75 000 美元，又 20%受过大学教育[②]。作为预选调查的一部分，将从这总体中抽取 1 600 人的简单随机样本。

   (a)为求样本中不超过 58%的人已婚的机会，需要一个盒子模型。盒中的票数应为 1 600 还是 100 000？解释之，并求出机会。

   (b)为求样本中不少于 11%的人年收入超过 75 000 美元的机会，需要一个盒子模型。盒子中的每张票子应该写上人们的收入吗？解释之。并求出机会。

   (c)求出样本中 19%和 21%之间的人受过大学教育的机会。
3. (a)第一节中的社会学家在她的容量为 100 的样本中男性在 41%和 51%之间的机会是多少？

   (b)在 25 组容量为 100 的样本中，约有多少组会出现男性人数在 41%和 51%之间？

   (c)在表 2 的第一行中，有多少个确实如此？（包括端点）。

(d)在表 1 的第一列中，所有元素为 F，是计算机出了毛病 吗？解释之。

4. 对容量为 400 的样本，重复习题 3(a)。

5. (a)第一节中的社会学家在容量为 100 的简单随机样本中恰好取得 46 名男性的机会是多少？

(b)在 50 个容量为 100 的样本中，约有多少个恰有 46 名男性？

(c)表 2 的前两行中有多少个恰好如此？

6. 从一只装有 1 颗红弹子和 9 颗蓝弹子的盒子中随机放回地抽取 900 次。样本中红弹子的百分数的 SE 为 1%。若样本百分数高于它的期望值 1SE，则它等于____。

$10\%+1\%$　　　　　$1.01\times10\%$

选择一个，并作简短解释。

7. 某人抛一枚硬币 10 000 次，并用“$\sqrt{10\ 000}\times0.5=50\%$”计算出现头像的百分数的 SE。这是正确的 SE 吗？回答是或不是，并作简短解释。

8. 盒子 | 0 | 0 | 0 | 1 | 2 | 的平均数为 0.6，SD 为 0.8。正确还是错误：在 400 次抽取中抽得 1 的百分数的 SE 可求出如下——

抽得 1 的个数的 $\text{SE}=\sqrt{400}\times0.8=16$

抽得 1 的百分率的 $\text{SE}=\frac{16}{400}\times100\%=4\%$

解释之。

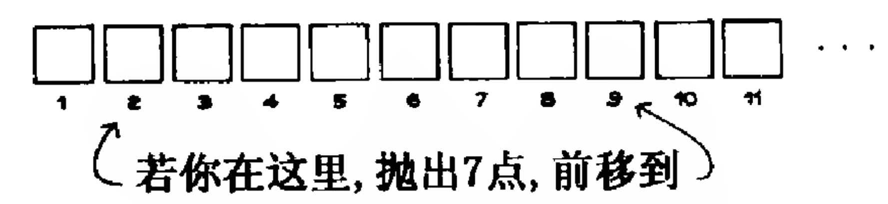
9. 某一所规模大的大学，全体注册学生的年龄分布未知。但是，在有 400 名学生的简单随机样本中，出现 200 人超过 21 岁，在下面答案中选择一个，并解释原因。

(i)恰好全体注册学生的 50% 超过 21 岁。

(ii)全体注册学生中约 50% 超过 21 岁，但可能偏离少许百分数点。

(iii)全体注册学生中约 50% 超过 21 岁，但可能偏离 10 或 20 个百分数点。

10. 某人做掷骰子游戏 100 次。每次他掷一对骰子，然后将他的标记沿一直线前移抛出点数之和个方格。(见图解)他们大约将可前移多远？加减多大误差？



11. (难题。)在某一总体中，身高的分布服从正态曲线，平均数为 5 英呎 8 英

时，标准差 SD 为 3 吋。从这总体中抽取容量为 500 的一个简单随机样本。为求样本中 7%和 11%之间的人身高超过 6 英吋的机会，需要一个盒子模型。盒中的票需要注上人的身高吗？解释之，并求出机会。

这些习题答案在 709—710 页。

技术性注释　当从 0—1 盒子中随机放回抽样时，抽得 1 的个数的 SE 为

$$\sqrt{\text{抽取次数}}\times\text{盒子的 SD}$$

故抽得 1 的百分数的 SE 为

$$(\sqrt{\text{抽取次数}}\times\text{盒子的 SD}/\text{抽取次数})\times 100\%$$

经代数运算，化简为（盒子的 SD/$\sqrt{\text{抽取次数}}$）×100%。在许多书中将它表成（$\sqrt{pq}/\sqrt{n}$）×100%，其中 p 是盒子中 1 的比率，q 是 0 的比率，n 是抽取次数。

3. 修正因子

1988 年劳动节刚过，Bush（布什）对 Dukakis（杜卡基斯）的总统竞选正在积极进行，焦点在西南。民意调查专家正试图预测出结果。在 New Mexico（新墨西哥）州有约一百二十万合法选民，在德克萨斯州有约一千二百五十万。假设一个民意调查机构在新墨西哥州取一 2 500 选民的简单附机样本，以便估计在该州的选民中民主党人的百分数。另一调查机构从德克萨斯州中取 2 500 选民的简单随机样本，以便估计那里民主党选民的百分数。两项民意调查使用完全相同的方法。每个估计由于机会误差都可能有点偏离。哪一个民意调查可能有较小的机会误差？

新墨西哥州的民意调查样本是从 500 选民中选 1 名，而德克萨斯州的民意调查是从 5 000 选民中选 1 名，似乎新墨西哥州的民意调查比德克萨斯州的民意调查精确。但是，这是直觉与统计理论正面冲突的地方之一，而且必须让步的正是直觉。事实上，从新墨西哥州所能期望的精度与从德克萨斯州所能期望的精度几乎相同。

当估计百分数时决定精度的是样本的绝对容量，而不是相应总体的大小，这在样本仅为总体的一小部分的通常情况下也是正确的。

盒子模型将有助于澄清这个问题。我们构造两只盒子，NM 和 TX。盒子 NM 代表新墨西哥州，盒子 TX 代表德克萨斯州。从盒子 NM 开始，它有 1 200 000 张票，每位选民一张，对应于民主党人的标 1，其余标 0。简单一些，我们假设 1 的百分比等于 50%。现雇用一家民意调查机构从盒子 NM 中取一简单随机样本，不告诉他们盒子中是什么。（别忘了，取一简单随机样本意指随机不放回抽样。）

民意调查机构的工作是估计盒中 1 的百分数，自然他们使用他们样本中 1 的百分数。他们不可能期望恰好正确。譬如，样本中 1 的百分数可能是 51%，它表明机会误差是 1 个百分数点。或样本百分数可能为 48%，对应的机会误差是 −2 个百分数点。一般，

$$\frac{\text{在取自盒子 NM 的 2 500}}{\text{张票子的样本中 1 的百分数}} = \frac{\text{盒子 NM 中}}{\text{1 的百分数}} + \text{机会误差}$$

对于盒子 TX，它是代表德克萨斯州的，因此有 12 500 000 张票，再次，我们将盒中的半数票子标 1，其余标 0。另一家民意调查机构从盒子 TX 中取一个 2 500 张票子的简单随机样本，同样不知道盒子的组成成分。他们也同样用样本中 1 的百分数估计盒子中 1 的百分数，且由于机会误差而有偏离，

$$\frac{\text{在取自盒子 TX 的 2 500}}{\text{张票子的样本中 1 的百分数}} = \frac{\text{盒子 TX 中}}{\text{1 的百分数}} + \text{机会误差}$$

盒子 NM 和盒子 TX 是按照相同的百分数组成成分构造的，且两个样本的容量相同。直觉将坚持认为从盒子 NM 抽样的机构将有较小的机会误差，因为盒子 NM 要小得多。但是统计学理论证明两项民意调查的机会误差的可能大小几乎相同。

这个问题现在可以更明确地陈述。统计学理论是怎样判断自己是正确的呢？首先，假设样本是按照放回方式抽取的。因此，用

哪一只盒子实际上毫无关系。在每次抽取中取得1和0的机会是50对50，与盒子的大小完全不相关。盒子NM和盒子TX有相同的SD即0.5，因此两家民意调查机构在它们的样本中1的个数的SE将相同：$\sqrt{2\,500}\times0.5=25$。从而两者的样本中1的百分数的SE也将相同：

$$\frac{25}{2\,500}\times100\%=1\%$$

因此，若他们随机放回地抽样，那么两家机构将偏离大约相同数值，1个百分数点左右，而与盒中容量的大小完全无关。

到现在为止，一切是令人满意的。但是支持这一论证的假设是抽样按照放回方式进行。而实际上，抽样是按照不放回方式进行的。不过，抽出的票子数目仅为盒子中票子数目的一个很小的部分。因此不放回地抽取票子几乎没有改变盒子中的组成成分。在每次抽取时，取到1的机会必定仍旧非常接近于50%，对0也类同，就机会而言，放回和不放回之间几乎没有差别。

本质上，那就是为什么总体的大小对估计的的精度几乎起不了什么作用的理由。不过放回和不放回抽样之间还是有一点区别的。对此统计学理论是如何考虑的呢？答案如下，当抽样是不放回时，盒子会变得略小点，略微减弱了变异性。因此不放回抽样的SE略小于放回抽样的SE。论述这问题的数学公式：

不放回抽样时的SE＝修正因子×放回抽样时的SE

有关这一因子的公式显得有点复杂：

$$\sqrt{\frac{\text{盒子中的票数}-\text{抽取的票数}}{\text{盒子中的票数}-1}}$$

当盒子中的票数相对于抽取的票数很大时，修正因子接近于1，可以忽略（表3）。因此，是样本的绝对容量通过放回抽样的SE来决定精度。总体的大小并不真正重要。（另一方面，若样本是总体的较大的一部分时，必须使用修正因子。）

表 3　修正因子。抽样次数固定为 2 500 次。

盒中的票数	修正因子(取 5 位小数)
5 000	0.07 718
10 000	0.86 609
100 000	0.98 743
500 000	0.99 750
1 000 000	0.99 875
12 500 000	0.99 990

在模型中，两只盒子的 1 的百分数是相同的。在实际场合，两个州的民主党人的百分数是不相同的。但是，差异一般不会大得要抛弃上述论点。譬如，在 1988 年的总统选举中，新墨西哥州 48% 的选民选择了民主党侯选人 Michael Dukakis(迈克尔·杜卡基斯)，在德克萨斯州是 44%，但是两个州的 SD 几乎相同：

48% [1] 52% [0]	44% [1] 56% [0]
NM	TX
$SD=\sqrt{.48\times.52}\approx.50$	$SD=\sqrt{.44\times.56}\approx.50$

容量为 2 500 的样本在德克萨斯州和在新墨西哥州将起到同样好的作用，尽管德克萨斯州要大十倍。下页漫画中的德克萨斯人是错的。

一个非数学的比拟将有所帮助。假设你从一只瓶中取一滴液体做化学分析。若液体是充分混合的，这一滴的化学成份会相当准确地反映整瓶的成份，根本不在乎这只瓶是一只试管或者是容量为一加仑的大壶，化学家不在乎这一滴是溶液的 1% 或 1% 的 $\frac{1}{100}$。

比拟是恰当的，对瓶中的每一分子，盒中有一张票与之对应，若液体是充分混合的，液滴是随机抽取。液滴中的分子数对应于抽取的样本中的票数。这数目——样本容量——是如此之大以致于百分数中的机会误差可忽略不计。



"一个大的州需要一个大的样本，伙计。"

习题 B

1. 一项民意测验所用简单随机样本的容量为 1 500，取自一个有 25 000 人口的城镇。另一项民意测验所用简单随机样本的容量为 1 500，取自 250 000 人口的城镇。其它条件相等：

   (i)第一项民意测验很可能比第二项民意测验有较高的精度。

   (ii)第二项民意测验很可能比第一项民意测验有较高的精度。

   (iii)这两项民意测验之间在精度上不可能有较大的差异。

   哪一种陈述是正确的？

2. 你雇了一家民意调查机构从一只有 100 000 张票的盒中取一简单随机样本，并估计盒中 1 的百分数。不告诉他们，盒中含 50%的 0 和 50%的 1。你该期望他们有多大的偏离：

   (a)若他们抽 2 500 张票？

   (b)若他们抽 25 000 张票？

   (c)若他们抽 100 000 张票？

3. 某调查机构要取一个简单随机样本以便估计收看某一电视节目的人的百分数。为了节约开支,他们想将样本取得尽可能小。但是他们的委托人只愿意接受估计中有 1 个百分数点左右的机会误差。他们应该使用容量为 100,2 500 或 10 000 的样本? 你可以假设总体非常大。
4. 设一盒中装有 2 颗红弹子和 8 颗蓝弹子。随机抽取 4 颗。求抽得红弹子的百分数的 SE,当抽样是

(a)放回　　　　　(b)不放回

这些习题的答案在第 711 页。

4. Gallup(盖洛普)民意测验

Gallup 民意测验从二亿合法选民中抽取几千人就以令人满意的精度预测选举的结果,这怎么成为可能的呢?前节的讨论集中于简单随机样本,但结论对抽取样本的大多数概率方法也是对的,包括 Gallup 民意测验所使用的那种:样本百分数中机会误差的可能大小主要依赖于样本的绝对容量,几乎完全不依赖于总体的大小。巨大数量的合法选民使抽取样本变得困难,但不影响标准误差。

2 500 是一个足够大的样本吗? 平方根法则提供了一个水准基点。例如,一枚硬币抛 2 500 次,头像的百分数的标准误差仅为 1%。类似地,对于 2 500 选民的样本,机会误差的可能大小亦仅为一个百分数点左右。一般地说,这是够好的,除非出现象 1960 年肯尼迪对尼克松那种选票非常接近的情况例外。选举团的情况比较复杂:Gallup 民意测验只预测公众选举。

5. 复习题

复习题可能包含前几章的内容。

1. 根据 Sherlock Holmos,

> 尽管单个的人是个不解之谜,但在整体中他变得数学上确实可知。例如,你不可能预言任何人将做什么,但你能精确地说出一个平均数量要做什么。单个的人是变化的,但百分数保持不变。统计学家这么说的[③]。

统计学家并不真如此说。Sherlock Holmos 忘了点什么?

2. 完成下面有关抛硬币游戏的表。

	头像次数		头像的百分数	
抛的次数	期望值	SE	期望值	SE
100	50	5	50%	5%
2 500				1%
10 000				
1 000 000				

3. 一只盒中装有 1 颗红弹子和 9 颗蓝弹子。从盒子中随机放回抽取弹子九百次。在抽取的弹子中红弹子的百分数应约为______,加或减______左右。填充并证明。

4. 一组合计 50 000 份的税收表格表明平均收入为 37 000 美元,SD 为 20 000 美元。此外,20%的表格表明总收入超过 50 000 美元。随机选取一组合计 900 份表格供审计。为了估计在选出供审计的表格中有 19%和 21%之间的人总收入超过 50 000 美元的机会,需要一个盒子模型。

(a)盒中的票子总数应为 900 还是 50 000?

(b)盒中的每张票子上应标明

0 或 1　　　　总收入

(c)正确或错误:盒子的 SD 为 20 000 美元。

(d)正确或错误:抽取数为 900。

(e)近似地求出在选出供审计的表格中有 19%到 21%的人总收入超过 50 000 美元的机会。

(f)根据已给的信息,你能(近似地)求出在选出供审计的表格中有 9%和 11%之间的人总收入超过 75 000 美元的机会吗?要么求出这个机会要么解释为什么你需要更多的信息。

5. 如在习题 4 中,只是希望求出(近似的)供审计表格的收入总和超过 33 000 000 美元的机会。回答(a)至(d)这部分问题;然后

求出机会或解释你为什么需要更多的信息。

6. 平均地说，旅馆的客人体重约150磅，SD为25磅。某工程师为一家普通旅馆设计一部载100人的大电梯。若她将电梯设计成可载15 500磅，对随机的100人一组它超载的机会最接近于

   1%的0.1　2%　5%　50%　95%　98%　99.9%

   选择一个，并解释之。

7. 普查局计划在每一个州取一个容量等于州人口1%的1/10的样本，以估计该州年收入超过50 000美元的州人口的百分数。其它条件相等：

   (i) 在加利福尼亚州(California)(人口三千万)所能期望的精度与在内华达州(Nevada)(人口一百万)所能期望的精度差不多相同。

   (ii) 在加州所能期望的精度比在内华达州要高得多。

   (iii) 在加州所能期望的精度比在内华达州要低得多。

   哪一种陈述是正确的？请解释之。

8. 一枚硬币抛400次，某人想用$\sqrt{400}\times 0.5=10$计算抛出头像次数的SE。这是正确的SE吗？回答是或不是，并简短地解释。

9. 某市有50 000户家庭，每户家庭中年满16岁或以上的平均人口已知为2.38；SD为1.87。某调查机构计划取一个400户家庭的简单随机样本，并访问在样本家庭中年满16岁或以上的全体成员。被访问的总人数约为____，加减____左右的误差。

10. 某镇有25 000户家庭，这些家庭平均拥有1.2部汽车；SD为0.90。又他们中的10%根本没有汽车。作为民意调查的一部分，选取了1 600户家庭的一个简单随机样本。试问9%和11%之间的样本家庭没有汽车的机会是多少？

11. 1965年，美国最高法院判决Swain V. Alabama的案子[④]。Swain是名黑人，在Alabama州Talladege郡被宣判犯有强奸一名白人妇女的罪。他被判死刑。案子上诉到最高法院，理由是在陪审团中没有黑人；甚至，没有黑人“在现仍活着的人的

记忆中曾在 Alabama 州 Talladege 郡的任何民事案件或刑事案件中出任过任何小陪审团成员。”

最高法院驳回上诉，理由如下。鉴于 Alabama 州法律规定，陪审团由近 100 人的陪审员名单中选取。陪审员名单中有 8 名黑人。（他们没有出任陪审团人员，因为检察当局以绝对反对方式将他们从名单中勾销。这种反对方式直到 1986 年都得到宪法保护。）在陪审团名单中有 8 名黑人表明“综合百分比的不一致性是小的，并说明不存在将一定数额的黑人纳入或排除之外的企图。”

那时在 Alabama 州，只有年过 21 岁的人有资格担任陪审团职务。在 Talladege 郡年过 21 岁的人有 16 000 人，其中近 26%是黑人。若从这总体中随机选取 100 人，其中黑人不多于 8 人的机会有多大？你得出什么结论？

6. 小结

1. 样本仅为总体的一部分，因此样本的百分比组成成分一般与整个总体的百分比组成成分少许不同。

2. 对于概率样本，机会误差的可能大小由标准误差给定。

3. 为求出百分数的 SE，需先求出相应数的 SE，然后换算成百分比。

4. 为了计算 SE，需要一个盒子模型。若问题包括分类和计数，或取百分比，则盒中只需有 0 和 1。如若必需时，可改变盒子。

5. 当样本仅为总体的一小部分时，在总体中个体的数目对估计 百分数的精度几乎没有影响。起作用的是样本的绝对容量（即样本中个体的数目），不是相对于总体的大小。

6. 当抽取是放回时，平方根法则是精确的。当抽取是不放回时，公式给出一个好的近似——只要抽取的票数与盒子中的票数相比是很小的话。

7. 当做不放回抽取时，要想获得精确的 SE 必须乘上修正因

子：

$$\sqrt{\frac{\text{盒中票数}-\text{抽取票数}}{\text{盒中票数}-1}}$$

8. 当盒中的票数远大于抽取的票数时，修正因子几乎等于 1。

21
百分数的准确性

在解答这类问题的过程中，重要的事是会倒推。那是一种非常有用的技能，并且还是非常容易的一种，但是人们常常不多去实践它……绝大多数人，如果你给他们描绘一系列事件，他们会告诉你结果会是什么。他们将这些事件都一起记在心中，并根据它们证明某件事将会发生。然而，只有少数人，如果你告诉他们某一结果，他们能根据自己的内在意识推出是哪些步骤导致这一结果。这种能力就是我说到倒推时所指的……

——Sherlock Holmos①

1. 引言

前一章是根据盒子对抽得数进行推理，抽得数是从组成成分已知的盒子中随机抽取的，典型的问题是求出抽得数中 1 的百分数落入某一给定区间的机会。象谢洛克·福尔摩斯指出的，将这种推理倒个向，即从抽得数到盒子，常常极为有用。统计学家将它称为根据样本推断总体，推断是本章的中心话题。

例如，假设某调查机构要知道某选区民主党人的百分数。他们可以取一个简单随机样本来估计它。自然，样本中民主党人的百分数可用来估计该选区民主党人的百分数——根据抽得数对盒子推理的一个例子。由于样本是随机选取的，因此只需根据样本的容量和成分就完全可能回答这估计可能有多准确。本章将说明这是怎么回事。

这种技巧是统计理论的关键思想之一，将在民意调查范畴内介绍它。某党派政治候选人，仅当他有获胜的良好机会时，才想在100 000合法选民的选区进入初选，他雇了一家调查机构，请机构取一个2 500选民的简单随机样本。在样本中，1 328人赞成该候选人，故百分数为

$$\frac{1\ 328}{2\ 500}\times 100\%\approx 53\%$$

该候选人正与他的民意调查人员讨论这一结果。

政治家：喏，我赢了。

民意调查者：没这么快。你需要知道在选区的全体选民中你能获得的百分数，我们只得到样本中的百分数。

政治家：但是有一个好的样本，必定是一样的。

民意调查者：不对。这就是我以前讲过的。你在样本中所获得的百分数不同于你在整个选区所能获得的。差异就是我们所说的机会误差。

政治家：样本会偏离多达三个百分点吗？如果这样，我可能失败。

民意调查者：事实上，你大约95%可以相信我们的调查正确到2个百分点之内，看来不错。

政治家：是什么告诉你机会误差的大小？

民意调查者：标准误差，别忘了，我们在另一天谈论过。就象我告诉你的……

政治家：对不起，我现在正在等一个电话。

政治家得出了在考虑调查数据时要问的关键问题：估计可能



“我支持你百分之 100，加或减 3 个百分点左右。”

会有多大的失误?正如民意调查者说的，机会误差的可能大小由标准误差给出，要计算它就需要一个盒子模型。

每一位选民在盒子中应该有一张票，总共是 100 000 张票。每张票上都标 1 或 0，这里 1 表示赞成该候选人，0 表示反对他。从盒子中随机抽取 2 500 张票。数据是抽得数，样本中赞成该候选人的选民人数等于抽得数之和。这完成了模型。

为了求出和的 SE，调查机构需要盒子的 SD。它是

$$\sqrt{(1\ 的比率)\times(0\ 的比率)}$$

在这一点上，他们似乎给难住了。他们肯定不知道盒中的每张票该怎么标。他们甚至并不知道盒子中 1 的比率。该参数表示选区中赞成该候选人的比率，他们被雇来正是为了求出它。

| ?? 个 $\boxed{0}$?? 个 $\boxed{1}$ |

100 000 张票

调查机构依靠他们自己的努力来克服这些障碍，他们用样本

中观察到的比率代替盒子中的未知比率。在本例中，2 500 人的样本里有 1 328 人赞成该候选人。故样本的$\frac{1\ 328}{2\ 500}=0.53$赞成他，另 0.47 反对他。估计为盒子中 100 000 张票的约 0.53 标 1，其余 0.47标 0。

在这基础上，盒子的 SD 的估计为$\sqrt{0.53\times0.47}\approx0.50$。于是，样本中赞成该候选人的选民人数的 SE 的估计为$\sqrt{2\ 500}\times0.50=25$。25 度量了 1 328 中的机会误差的可能大小。而 2 500（样本容量）中的 25 人为 1%。故样本中赞成该候选人的选民百分数的 SE 的估计为 1 个百分数点。这完成了估计标准误差的自助程序。

就该候选人来说，上述计算表明他的民意调查人员所得的估计 53%仅可能有 1 个百分点左右的偏离。极少可能有 3 个百分点的偏离——即 3 个 SE。因此，他完全位于 50%的安全一边。他应该进入初选。

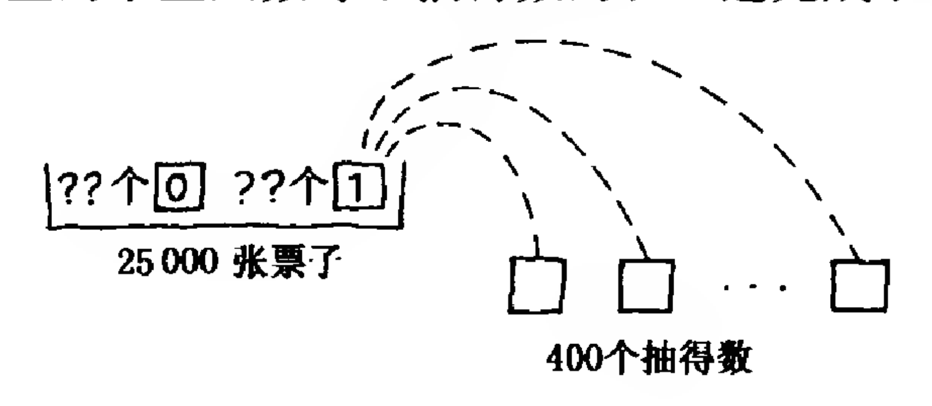
> 当从成分未知的 0—1 盒子中抽样时，可以用样本中 0 和 1 的比率代替盒子中未知的比率以便估计盒子的 SD。如果样本充分大的话，这个估计是好的。

上述自助程序似乎有点粗糙。但是，即使用中等容量的样本，抽得数中 1 的比率可能相当接近盒子中的比率，对 0 也一样。因此，如果调查机构在公式中用样本比率代替盒子的 SD，在估计它们的 SE 时他们不大可能有大的失误。

例 1. 1987 年秋，某市立大学有 25 000 名注册学生。那学期进行了一项调查以估计住在家里的学生的百分数，抽取了一个 400 名学生的简单随机样本，结果是 317 名住在家里。试估计那学期该大学住在家里的学生的百分数，且附带求估计的标准误差。

解. 样本百分数为(317/400)×100%≈79%，它是总体百分

数的估计。至于标准误差，需要一个盒子模型。盒子中有 25 000 张票，每张对应总体中的一名学生。从盒子中抽取 400 张，每张对应样本中的一名学生。这问题包括了分类和计数，因此盒子中的每张票应当标 1 或 0。我们对住在家里的学生计数，对应这些学生的票应该标 1，其余标 0。从盒子中随机抽取 400 张。数据是抽得数，样本中住在家里的学生人数等于抽得数的和。这完成了模型。



盒子中 1 的比率是参数。它表示 1987 年秋这所大学住在家里的全体学生的比率。它是未知的，可估计为 0.79——样本中这类学生的比率。类似地，盒子中 0 的比率估计为 0.21。故根据自助法，盒子的 SD 可估计为$\sqrt{0.79\times0.21}\approx0.41$。于是，样本中住在家里的学生人数的 SE 的估计为$\sqrt{400}\times0.41\approx8$。这是样本容量 400 的 2%。故样本百分数的 SE 的估计为 2%。从而，总体百分数约为 79%，约 2 个百分数点左右的误差。这是答案，约有 79%(±2%)的学生住在家里。

本节的讨论集中于简单随机抽样，这样数学上最简单。在实际应用中，调查机构使用复杂得多的设计。即便如此，利用概率方法，总可以说出一个估计可能有多大的失误。这是用概率方法抽取样本的巨大优越性之一。

习题 A

1. 本题是关于本节中的例 1 的。在样本中，317 名学生住在家里，在 317 中机会误差可能多大?
2. 一盒中装有 100 000 颗弹子，一些是红的其余是蓝的。盒中红弹子的百分数未知，为了估计它，抽取了 1 600 颗弹子的简单随机样本，样本中 322 颗

弹子是红的。盒子中红弹子的百分数的估计为______，±____左右。

3. 某市有 100 000 人年龄在 18 到 24 岁。从这年龄段的人中抽取一个 500 人的简单随机样本，其中 194 人最近刚入高校。估计该市最近入高校的年龄至 18 到 24 岁的全体人员的百分数[②]。并在估计上标出±数值。

4. 取自某学院毕业生的一个 100 人的简单随机样本中，48 人的年收入不少于 50 000 美元。估计该学院毕业生中年收入不少于 50 000 美元的所有毕业生的百分数[③]。附带估计标准误差。

5. 从某州所有制造企业的总体中抽取一个容量为 400 的简单随机样本：样本中有 16 家企业的雇员不少于 250 人，估计雇员不少于 250 人的制造企业的百分数[④]。附带估计标准误差。

6. 在同一个州中，从受雇于制造企业的所有人的总体中取一个容量为 400 的简单随机样本：样本中 216 人在雇员不少于 250 人的企业中工作。试估计在雇员不少于 250 人的企业中工作的人的百分数。附带求出估计的标准误差。

7. 习题 5 和 6 中的两个百分数间的差异应归因于机会误差吗？

下面二道习题是为了说明估计盒子的 SD 的(自助)bootstrap 法而设计的。

8. 假设盒子中有 100 000 张票，每张标 0 或 1。设盒中实际上有 20%的票是 1。试计算从盒中抽取的 400 张票子中 1 的百分数的标准误差。

9. 三个人从习题 8 的盒子中各取一个容量为 400 的简单随机样本，他们不知道盒中内容。第一个样本中 1 的张数是 72；第二个是 84；第三个是 98。每人都用 bootstrap 方法估计 SE。

(a)第一人估计盒中 1 的百分数为____，±____左右。

(b)第二人估计盒中 1 的百分数为____，±____左右。

(c)第三人估计盒中 1 的百分数为____，±____左右。

这些习题的答案见第 711—712 页。

2. 置信区间

在上节的例子中，样本里有 79%的学生住在家里：样本的百分数为 79%。总体的百分数与 79%的差有多大？(别忘了，“总体的百分数”意指该大学全体住在家里的学生的百分数。)

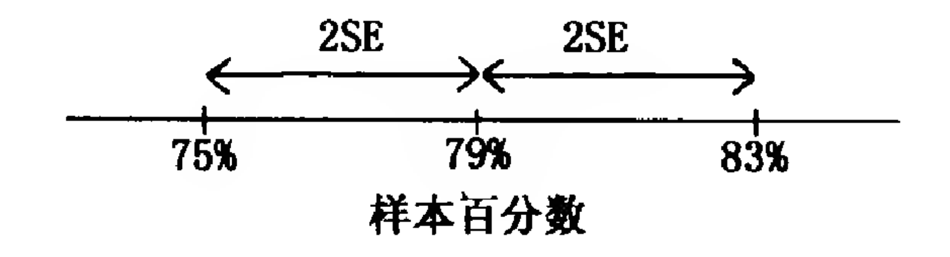
估计标准误差为2%，表明机会误差的大小约为2%。故总体的百分数很可能为77%。意思是2%的机会误差：

样本的百分数＝总体的百分数＋机会误差

79%＝77%＋2%

总体的百分数也可能是76%，对应3%的误差。这变得不那么可能，因为3%代表1.5个SE。总体的百分数甚至可能小到75%，但这就更加不可能；4%的误差代表2个SE。当然，总体的百分数可能落在样本的百分数的另一边，对应着负的机会误差。譬如，总体的百分数可能是83%。从而估计低了4%，机会误差是－4%，即－2个SE。

由于机会误差，在可能与不可能之间没有明显界限。大于2个SE的误差会发生——不太经常。若你在2个SE处切断会出现什么呢？换句话说，取一个由样本的百分数以下2个SE到以上2个SE的区间：



这是总体的百分数置信水平约为95%的置信区间。你可能有约95%的把握在区间75%到83%内逮住总体的百分数。

如果你需要不同的置信水平又会怎样呢？除了100%之外都是可能的，只要从样本的百分数向两边各移离适当数目的SE。譬如：

- 区间“样本的百分数±1个SE”是总体的百分数的近似68%置信区间。
- 区间“样本的百分数±2个SE”是总体的百分数的近似95%置信区间。
- 区间“样本的百分数±3个SE”是总体的百分数

的近似 99.7%置信区间。

没有 SE* 的倍数能给出 100%的把握，因为总存在极小的可能性会出现非常大的机会误差。数学上这一点由正态曲线没有限定的范围这一事实所反映。无论你选的有穷区间有多大，正态曲线总有某一部分面积在该区间之外。

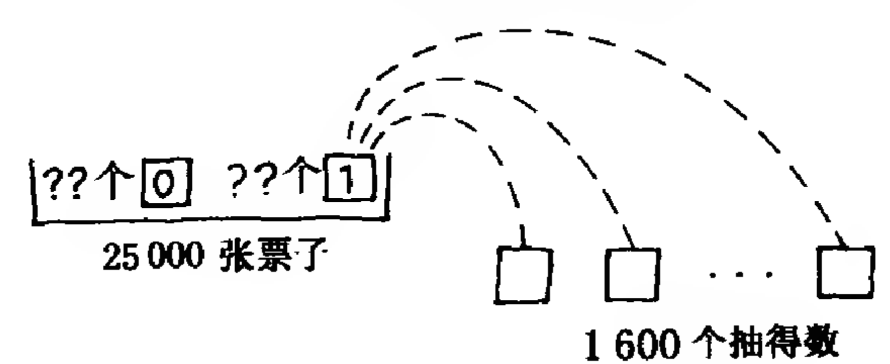
例 2. 取一个 1 600 人的简单随机样本来估计某镇 25 000 名合法选民中民主党人所占百分数。结果为样本中 917 人是民主党人。试求全体 25 000 名合法选民中民主党人所占百分数近似的 95%置信区间。

解　样本中民主党人的百分数为

$$\frac{917}{1\ 600}\times 100\%\approx 57\%$$

估计值为约 57%的合法选民是民主党人。

至于标准误差，需要一个盒子模型。对镇上的每一位选民，盒中有一张票与之对应，构成总共 25 000 张票。抽取 1 600 张对应样本容量 1 600。这问题包括了选民的分类(是民主党人或不是)和计数，故每张票上标 1 或 0。被计数的是民主党人，故对应于民主党人的票子标 1，其余标 0。从盒中随机抽取 1 600 张。数据就是抽得数，样本中民主党人的人数就是抽得数之和。这完成了模型。



盒子中 1 的比率(译意:25 000 合法选民中民主党人的比率)是未知的，但是可以用样本中民主党人的比率 0.57 来估计。类似地，可估计盒子中 0 的比率为 0.43。故盒子的 SD 可以用自助法估计为$\sqrt{0.57\times 0.43}\approx 0.50$。从而可将样本中民主党人的人数的 SE 估计为$\sqrt{1\ 600}\times 0.50=20$。化成样本容量的百分数，这等于

$$\frac{20}{1\ 600}\times 100\%=1.25\%$$

因此样本中民主党人的百分数的SE为1.25%。换句话说，样本中民主党人的百分数可能偏离全体25 000合法选民中民主党人的百分数1.25个百分数点左右。在全体合法选民中民主党人所占百分数的一个近似95%置信区间为

$$57\%\pm 2\times 1.25\%$$

这就是答案。我们约95%可以相信这个镇上25 000名合法选民中的54.5%到59.5%之间为民主党人。

置信水平被引述为“大约”这么一种程度。一个理由是标准误差已经由数据估计出。另一个理由是使用了正态近似。若不运用正态近似，本章的方法就不能实施。不存在严格的判定规则。进行的最好方式：设想总体与样本有相同的百分比成分。然后，试图判定正态近似能否适用于取自对应盒子的抽得数之和。

例如：样本的百分数接近0%或100%表明盒子的成分偏向一边，而且接受正态近似之前需要进行大量的抽取（第18章，第5节）。另一方面，如果样本的百分数接近50%，只要有100次左右的抽取，正态近似应当是满意的。

习题B

1. 本题是关于本节例2的。在样本中，有917名民主党人，在917中机会误差的可能大小有多大？
2. 重新参阅上节习题A中习题2。
   (a)试求盒中红弹子的百分数的近似95%置信区间。
   (b)若可能，对置信水平99.7%再求一次。
3. 重新参阅上节习题A的习题3。试求该市所有年龄在18到24岁中最近入高校的人的百分数的近似95%置信区间。
4. 如题3，但设样本容量为2 500，他们中972人最近入高校。
5. 一只盒子装有1颗红弹子和99颗蓝弹子；从盒中随机放回地抽取100颗弹子。

(a)求抽出的弹子中红弹子数的期望值，和 SE。

(b)抽得少于 0 颗红弹子的机会是多少？

(c)用正态曲线估计这个机会。

(d)抽出弹子中红弹子数的概率直方图看起来象正态曲线吗？

6. 一只盒子中装有 10 000 颗弹子，其中有些是红的，其余是蓝的。为了估计盒中红弹子的百分数，随机不放回地抽取 100 颗弹子。在抽出的弹子中有 1 颗是红的。盒中红弹子的百分数估计为 1%，具有 SE 为 1%。正确还是错误：盒中红弹子的百分数的近似 95% 置信区间为 1%±2%，解释之。

这些习题的答案在第 712—713 页。

3. 置信区间的注释

回到例 1，抽取一个简单随机样本以估计 1987 年秋季某大学当时住在家里的注册学生的百分数。这百分数的一个近似 95% 置信区间是从 75% 到 83%，因为

$$样本的百分数 \pm 2SE = 75\% 至 83\%$$

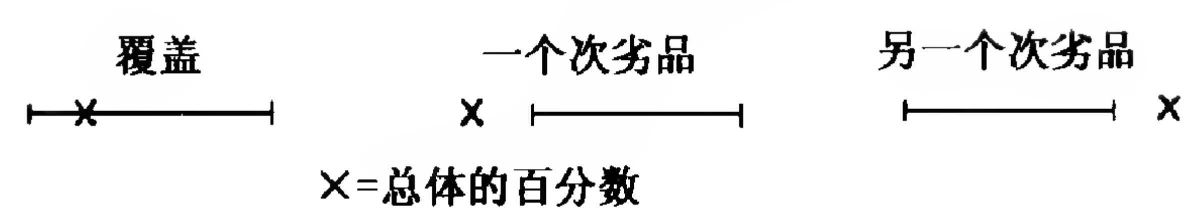
说成“总体的百分数在 75% 和 83% 之间有 95% 的机会”似乎更自然些。但是这里有一个问题。在频率理论中，机会表示某事物发生次数的百分数。无论你对该校 1987 年秋注册的全体学生估量多少次，当时住在家里的学生的百分数不会改变的。或者这百分数在 75% 和 83% 之间，或者不在。因此，实际上无法确定该百分数落在区间 75% 到 83% 之内，或落在其它区间的机会。这就是为什么统计学家要把问题少许变通一下。他们认识到机会存在于抽样过程中，不是在参数里。他们用一个新词“置信”提醒你注意这一点。

机会存在于抽样过程中，不在参数里

95% 的置信水平只能说明有关抽样过程中的某件事，我们来看一下那是什么。要注意的第一点：置信区间依赖于样本。如果样本不同，置信区间也就不同。对某些样本来说，区间“样本的百分数 ±2SE”抓住了总体的百分数。（统计学家用覆盖一词。）但是对另一些样本，区间未能覆盖，就象购买一辆旧汽车那样，有时你买到

的是次劣品——一个未能覆盖参数的区间。

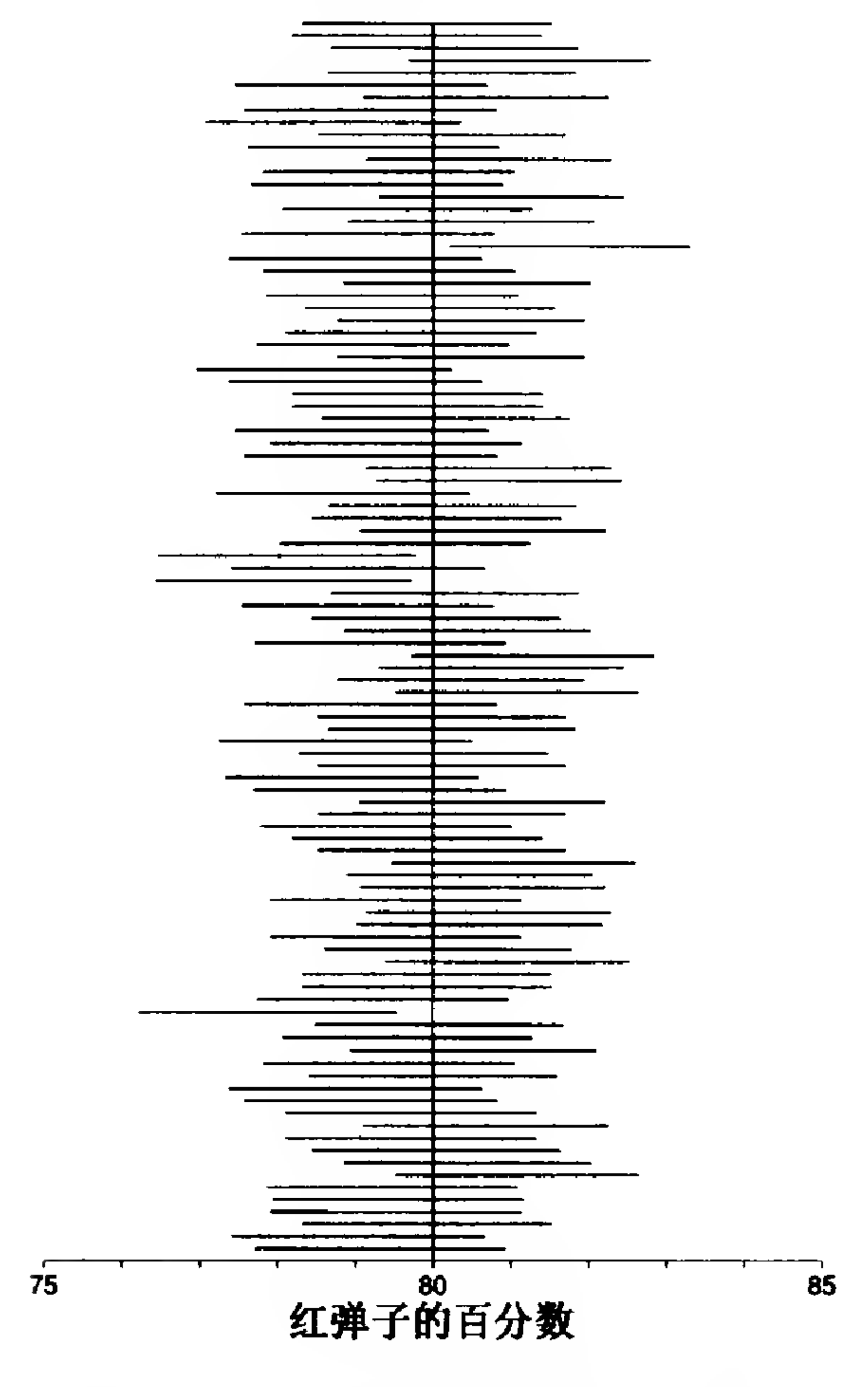
三个置信区间



区间“样本的百分数±2SE”有95%的置信水平。现在可以对它作出注释了。所有次数抽样样本中的大约95%，区间“样本的百分数±2SE”能覆盖总体的百分数，而另外的5%的次数则不能覆盖。当然，调查人员一般不可能宣称他们的某一特定区间是否覆盖了总体的百分数。他们不知道该参数，它就是他们想要估计的东西。但是有一种程序，它95%的时候能做到覆盖该参数：取一个简单随机样本，然后从样本的百分数向两边移出2个SE。就这程序来说，它就象从一只装满区间的盒子中随机抽取区间，而盒子中所装的区间有95%是覆盖参数的，仅5%是次劣品。它描述购买旧汽车的问题。

置信水平的这一注释略为难些，因为它包括了不仅考虑事实上的样本，而且也要考虑其它可能抽取的样本。注释在图1中描述。一百家调查机构受雇估计一只大盒子中红弹子的百分数。这个百分数是80%，而调查机构是不知道的。每家机构取一个2 500颗弹子的简单随机样本，并用公式“样本中红弹子的百分数±2SE”，计算出盒中红弹子的百分数的一个95%置信区间。红弹子的百分数由于不同的样本而不同，标准误差的估计亦是如此。结果是这些置信区间有不同的中心和长度。某些调查机构得到覆盖盒中红弹子的百分数的区间，另一些则失败了。它们中有约95%应该覆盖盒中红弹子的百分数，这个百分数由一条垂直线标出。实际情况是，100个中有96个如此[5]。

图 1　置信区间的注释。对 100 个不同样本证实 95%置信区间，区间随样本变化而变化。约 95%的样本，其区间覆盖由一条垂直线标出的总体的百分数[⑥]。



这里有许多要考虑，但是记住本章的中心思想：

由于机会误差，样本的百分数将偏离总体的百分数。SE 将告诉你偏离数值的可能大小。

引入置信水平是为了使这一思想更定量化。

习题 C

1．参阅本章习题 A 中的习题 8 和 9。试用习题 9(a)中的人所得数据算出盒

子中 1 的百分数的 95%置信区间。对另外两人的数据重做一遍。三个区间中哪些个覆盖总体的百分数，即盒子中 1 的百分数？哪些个未能覆盖？（记住，题 9 中的三个人并不知道盒中所装的内容；但你从习题 8 中知道。）

2. 一枚硬币抛 100 次，抛出头像的百分数的期望值为 50%，SE 为 5%。说出下列各陈述是正确的还是错误的，并简短地解释之。

(a)5%度量了 50%中机会误差的可能大小。

(b)在抛 100 次的结果中头像的百分数约为 50%，±5%。

(c)在抛 100 次的结果中头像的百分数的近似 95%置信区间为从 40%到 60%。

3. 一盒中装有大量的红弹子和蓝弹子，但比例未知；随机模出 100 颗弹子，有 53 颗是红的。盒子中红弹子的百分数估计为 53%，SE 算出为 5%。说出下列各陈述是否正确，并简短地解释之。

(a)5%度量了 53%中机会误差的可能大小。

(b)53%可能偏离盒子中红弹子的百分数 5%左右。

(c)盒子中红弹子的百分数的近似 95%置信区间为从 43%到 63%。

(d)样本中红弹子的百分数的近似 95%置信区间为从 43%到 63%。

4. 抽取一个 1 000 人的简单随机样本以估计一个大的人口总体中民主党人的百分数。样本中有 543 人是民主党人。样本的百分数为(543/1 000)×100%=54.3%。样本中民主党人的百分数的 SE 计算为 1.6%。正确还是错误，并解释之：

(a)54.3%±3.2%是总体百分数的 95%置信区间。

(b)54.3%±3.2%是样本百分数的 95%置信区间。

(c)对总体中民主党人的百分数，三分之二的机会落在 54.3%±1.6%中。

5. （续习题 4）若另一家调查机构取一个 1 000 人的简单随机样本，在他们的样本中民主党人的百分数约有 99.7%的机会将在范围 54.3%±4.8%中。

6. 在某所规模较大的大学里，54%的学生是女姓，46%是男性。从这总体中抽取一个 1 000 人的简单随机样本。样本中女性百分数的 SE 计算为 1.6%。正确还是错误：样本中女性的百分数约有三分之二的机会落在范围 54.3%±1.6%中。解释之。

这些习题的答案参见第 713 页。

4. 货物出门概不退换

本章的各种方法是对简单随机样本产生的。它们不适用于其

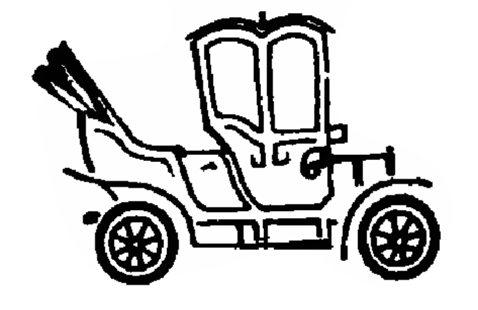
它类型的样本。大多数调查机构使用相当复杂的概率方法抽取他们的样本(第 19 章的第 4 节)。其结果是,他们必须用较复杂的方法估计他们的标准误差。有些调查机构根本就不使用概率方法。注意提防他们。

> 告诫,有关简单随机样本的公式不可用于其它类型的样本。

理由在于:逻辑上,本章的各种程序全都出自平方根法则(第 17 章的第 2 节)。当样本的容量相对于总体的容量很小时,抽取一个简单随机样本几乎与从盒子中随机放回地提取相同——平方根法则适用的情况。这里"随机"一词是指其专门含义:在每一阶段,盒子中的每张票有相等的机会被抽取。若样本不是随机地取,平方根法则就不适用,且可能导致愚蠢的答案⑦。

人们常常认为统计公式在使用时将以某种方式检验自己,以确定它应用了。没有什么能远离真理。在统计学中,就象在老式资本主义一样,责任在消费者。

货物出门概不退换　　　　买者谨访



$$\bar{x} \pm z_\alpha \times s/\sqrt{n}$$

习题 D

1. 某心理学家对他班上的 100 名学生施行一项被动性测验,并发现其中的 20 人得分超过 50。他断定近似地全部学生的 20% 对这项测验得分会超过 50。他承认这估计可能有少许偏离,并估计误差的可能大小如下:

$$\text{人数的 SE} = \sqrt{100} \times \sqrt{0.2 \times 0.8} = 4$$

$$\text{百分比的 SE} = \frac{4}{100} \times 100\% = 4\%$$

统计理论说明什么?

2. 一所规模较小只招收本科生的学院有 1 000 名学生，平均分布在四个班级：一年级，二年级，三年级和四年级。为了估计吸过大麻的学生的百分数，用下面程序取了一个样本：随机不放回地从四个班级的每一个中取 25 名学生。结果是，100 名样本学生中 35 人承认吸过。因此，估计该学院的 1 000 名学生 35％会承认吸食过大麻。隶属于这一估计的标准误差由下面程序给出：

$$\text{人数的 SE}=\sqrt{100}\times\sqrt{0.35\times0.65}\approx5$$

$$\text{百分比的 SE}=\frac{5}{100}\times100\%=5\%$$

统计理论说明什么？

这些习题的答案在第 714 页。

5. Gallup(盖洛普)民意测验

盖洛普民意测验不用简单随机样本(第 19 章的第 4 节)。因此，他们不用本章的方法估计他们的标准误差。但是，把他们的样本与相同容量的简单随机样本做比较是有趣的。例如，1952 年他们基于一个 5 385 人的样本，预测 51％的选票赞成艾森豪威尔。完成对简单随机样本的计算，

$$\text{人数的 SE}=\sqrt{5\ 385}\times\sqrt{0.51\times0.49}\approx37$$

$$\text{百分比的 SE}=\frac{37}{5\ 385}\times100\%\approx1\%\text{的 }0.7$$

实际上，在那次选举中艾森豪威尔获得选票的 55.4％。盖洛普民意测验的估计偏离了 4.4 个百分数点，它比简单随机样本的 SE 的 6 倍还大。表 1 列出 1952 年至 1988 年历届总统选举的比较情况。

表 1 盖洛普民意测验与简单随机样本的比较。总的来看，预测误差比相同容量的简单随机样本所期望的误差要大得多。

年份	样本容量	简单随机样本的 SE	实际误差
1952	5 385	1%的 0.7	4.4%
1956	8 144	1%的 0.6	1.7%
1960	8 015	1%的 0.6	1%的 0.9
1964	6 625	1%的 0.6	2.7%
1968	4 414	1%的 0.8	1%的 0.5
1972	3 689	1%的 0.8	1%的 0.2
1976	3 439	1%的 0.9	1.6%
1980	3 500	1%的 0.8	3.7%
1984	3 456	1%的 0.9	1%的 0.2
1988	4 089	1%的 0.8	2.1%

来源：见第 19 章的表 4。

表 2 盖洛普民意测验与其容量等于盖洛普民意测验中的可能参加投票的人数的简单随机样本之间准确性的比较。

年份	可能参加投票人数	简单随机样本的 SE	实际误差
1952	3 350	1%的 0.9	4.4%
1956	4 950	1%的 0.7	1.7%
1960	5 100	1%的 0.7	1%的 0.9
1964	4 100	1%的 0.8	2.7%
1968	2 700	1.0%	1%的 0.5
1972	2 100	1.1%	1%的 0.2
1976	2 000	1.1%	1.6%
1980	2 000	1.1%	3.7%
1984	2 000	1.1%	1%的 0.2
1988	2 600	1.0%	2.1%

注：可能参加投票选民人数是以整数来表示的。

来源：盖洛普民意测验（美国民意协会）。

在 10 次选举中有 7 次盖洛普民意测验的误差比对简单随机样本算出的 SE 大得多。一个理由是，盖洛普民意测验的预测仅基于一部分样本，即被判定可能参加投票的那些人（第 19 章的第 6 节）。这排除了约一半的样本。表 2 把盖洛普民意测验产生的误差

与对简单随机样本，其容量等于民意测验中所访问的可能参加投票的选民人数，所算出的SE相比较。简单随机样本的公式在预测误差的大小上仍然效果不好。这是由于盖洛普民意测验取的不是简单随机样本。

习题 E

1. 在某次选举中，盖洛普民意测验预选调查基于一个1 000人的样本估计65%的选票赞成民主党候选人。正确还是错误，并解释之：在这估计中机会误差的可能大小可以计算如下——

$$\sqrt{1\ 000}\times\sqrt{0.65\times0.35}\approx15 \qquad \frac{15}{1\ 000}\times100\%=1.5\%$$

2. 从一只大盒子中随机不放回地抽取1 000张票，抽出的票中651张为1，盒中标1的比率估计为65%。正确还是错误，并解释之：这估计中机会误差的可能大小可以计算如下——

$$\sqrt{1\ 000}\times\sqrt{0.65\times0.35}\approx15 \qquad \frac{15}{1\ 000}\times100\%=1.5\%$$

3. 下面的文章发表在《纽约时报》，1988年8月27日的Op Ed专页上，标题是"可能布什已经获胜"。

   总统竞选仅仅现在才正式大张旗鼓地开始，其实，已经实质上结束。尽管如Niagara（尼亚加拉）瀑布一样的有关于候选人如何招徕他们的竞选伙伴，如何通过辩论不休以及相互诽谤中伤等等的新闻故事，但是木已近于成舟。

   一个有效的指示器是盖洛普民意测验，它本周指出副总统布什领先Michael S · Tukakis（迈克尔 · S · 杜卡基斯）4个百分数点。自从George Gallup开始他的在总统选举年的选民意向调查以来的半个世纪中，他在9月份最后一个星期左右的"预赛"都以显著的准确性预告了选举的结果。

   已故的James A · Farley，50年前民主党无可匹敌的战术家，常争辩说选民到劳动节左右时拿定他们的主意。……此外，现已证实，如果当传统不参加投票的人——抵制投票宣传的对象——被说服参加投票的话，他们也会遵循全体选民中其他人的相同比例投他们的票……9月份到11月份百分数的显著改变只能归因于被改变了的选民热情。

(a)别的还有什么能解释9月底盖洛普民意测验的结果与11月初的选举结果之间的差异？

(b)9月底盖洛普民意测验的结果与11月初的选举结果之间有几个百分数点的差别是：极其不可能的，未必但或多或少可能，非常可能。选其中一种见解，并解释之。

这些习题的答案在第714页。

6. 复习题

复习题可能包含前几章的内容。

1. 住宅能源消耗调查发现在1984年美国家庭的37.6%拥有洗碟机，99.7%拥有电冰箱[8]。某市场调查机构在一个有25 000户家庭的镇上重复这项研究，采用一个500户家庭的简单随机样本：样本家庭中的179户拥有洗碟机，498户拥有电冰箱。

   (a)该镇拥有洗碟机的家庭的百分数估计为____，这估计可能偏离____左右。

   (b)如若可能，求出25 000户家庭中拥有洗碟机的百分数的95%置信区间。如若不可能，解释为什么不行？

   (c)对电冰箱重复(a)部分。

   (b)对电冰箱重复(b)部分。

2. 续习题1，在样本家庭中，121户不拥有汽车，172户拥有一辆汽车，207户拥有两辆或更多汽车。试估计该镇拥有汽车的家庭的百分数；附带求出估计的标准误差。或这可由给定信息求出吗？

3. 1986年，全国教育进展评审会对一个在校17岁青少年的全国性样本考核历史和文学。你可以假设取了一个容量为6 000人的简单随机样本[9]。仅36.1%知道(Chaucer)乔叟写作《坎特伯雷故事集》(The Canterbary Tales)，但95.2%知道Edison(爱迪生)发明电灯泡。

   (a)如若可能，求出全体在校17岁青少年中知道乔叟写作《坎特伯雷故事集》的百分数的近似95%置信区间。如若这是

不可能的，为什么不呢?

(b) 如若可能，求出全体在校 17 岁青少年中知道爱迪生发明电灯泡的百分数的 95% 置信区间。如若这是不可能的，为什么不呢?

4. 从某一大城镇中抽取一个 3 500 人的年龄不小于 18 岁的简单随机样本以估计(该镇年龄不小于 18 岁)阅读报纸的人的百分数。样本中有 2 487 人是报纸读者[10]。总体的百分数估计为

$$\frac{2\ 487}{3\ 500}\times 100\%\approx 71\%$$

标准误差估计为 1% 的 0.8，因为

$$\sqrt{3\ 500}\times\sqrt{0.71\times 0.29}\approx 27 \qquad \frac{27}{3\ 500}\times 100\%\approx 1\% \text{的 } 0.8。$$

1% 的 0.8 是正确的加或减误差吗?回答是或不是，并解释之。

5. (假设的)某年，纽约证券交易所有 252 个交易日，IBM 的普通股在其中的 131 天是上升的：$\frac{131}{252}\times 100\%\approx 52\%$。某统计员在这估计上附加标准误差为下：

$$\text{天数的 SE}=\sqrt{252}\times\sqrt{0.52\times 0.48}\approx 8$$

$$\text{百分比的 SE}=\frac{8}{252}\times 100\%\approx 3\%$$

这是一个正确的 SE 吗? 回答是或不是，并解释之。

6. (假设的)在某一城市一家银行想要估计人们携带零钱的数额。他们取了一个 100 人的简单随机样本，并发现平均来说，样本中的人携带 73 美分的零钱。他们计算得标准误差为 4 美分，因为

$$\sqrt{100}\times\sqrt{0.73\times 0.27}\approx 4 \qquad \frac{4}{100}\approx 0.04$$

他们是正确的吗? 回答是或不是，并解释之。

7. 正确还是错误：对一个设计良好的抽样调查，样本的百分数是非常可能等于总体的百分数的。解释之。

8. 在以 Keno 命名的游戏中，有 80 只编号从 1 到 80 的球，从中随机抽取 20 只。若你赌某一对数，当这两个数都被抽出时你赢。赌

注是 1 赔 11，你有非常接近于 6% 的赢的机会[11]。如果你赌 100 次，每次在某一对数上押 1 美元，你的净收益约为____±____左右。

9. 从一只装有编号票子的大盒子中随机不放回地抽取 100 张。有两个选择：

(i)若抽得数之和大于 710 则赢 1 美元。

(ii)若抽得数的平均数大于 7.1 则赢 1 美元。

哪一个有利些？或它们是相同的？解释之。

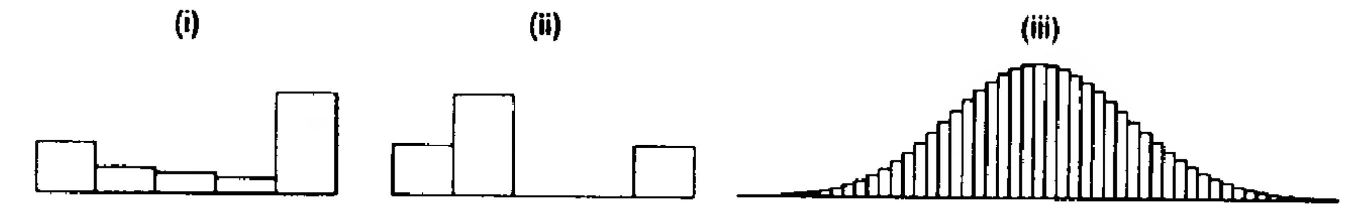
10. 某项每月一次的民意调查是基于一个 1 500 人的样本，“经科学选取作为美国公众的一个有代表性的剖面”。新闻界发布告诫说这些估计值都含有机会误差，但保证他们“在两个百分数点之内是可靠的”。“可靠的”一词意义不明确。根据统计理论，该保证应述说如下：

(i)事实上所有这些调查中，估计值会落在参数的两个百分数点之内。

(ii)在大多数的这种调查中，估计值会落在参数的两个百分数点之内，但是在某确定的百分比的次数中预期有较大的误差。

解释之。

11. 从盒子 [1] [2] [2] [5] 中随机放回地抽取 100 次。下图中，有一张是抽得数个数的直方图。另一张是抽得数之和的概率直方图。而第三张是无关的。哪一张对应哪一种直方图？为什么？



7. 小结

1. 要求估计一个大的总体中具有某种给定特性的人的百分

数。这个参数称为总体百分数，它是未知的。取一个简单随机样本。样本中具有这种给定特性的人的百分数是一个统计量，称为样本百分数。样本百分数被用来估计总体百分数。

2. 由于机会误差，样本百分数会偏离总体百分数。样本百分数的 SE 可以告知你偏离值的可能大小。

3. 当从成分未知的 0—1 盒子中抽样时，盒子的 SD 可以通过用样本中 0 和 1 的比率代替盒子中未知的比率而得到估计。当样本足够大时，自助估计是好的。计算 SE 需要用到 SD。

4. 总体的百分数的置信区间可通过从样本的百分数向两边移离合适数目的标准误差获得。置信水平可从正态曲线读出。这方法只能对大样本使用。

5. 在概率论的频率理论中，参数不受机会变化的支配，这就是为什么作出置信陈述来取代概率陈述。

6. 简单随机样本的有关公式不可用于其它类型的样本，若样本不是依概率方法选取，务必当心。

22
估量就业与失业

国家渴望信息；统计特征的，或即使有统计面孔的每件事都以一种近乎可怜的渴望心情被接受；公众还没有学会对这类陈述有所需的一半的怀疑和批判。

——Francis A · Walker 将军，1870 人口普查负责人

1. 引 言

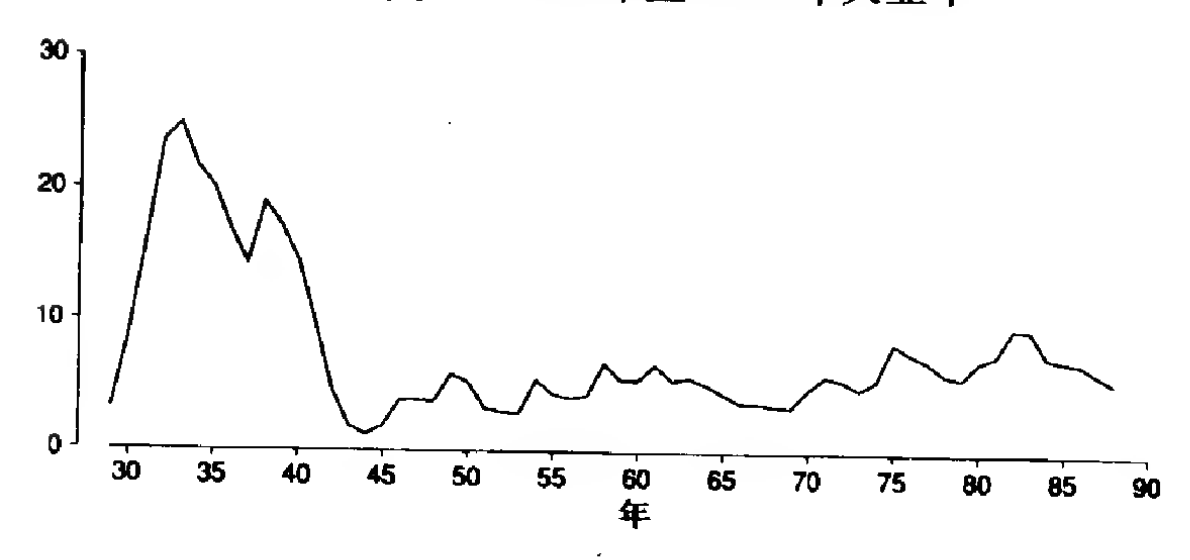
失业率是政府发布的最主要经济指标之一。它的曲线如图 1 所示。1929 年，证券市场崩溃之前，失业率仅为 3%。在大萧条的深渊中达到 25%，并维持相当地高直到美国卷入第二次世界大战。近些年来，作为联邦储备部门 1981 年采取反通货膨胀措施的结果，1982—1983 年经济深陷低谷，失业率达到近 10%。到了 80 年代末期，失业率再次下降到 6%；且在许多大城市范围内存在着技术工人短缺现象。

主管估量就业和失业状态的政府机构是劳工统计局。但是，他们是怎么知道谁受雇用或未受雇用的呢?实际上，各种就业统计是根据样本调查(现场人口调查)估计的。这种大规模和组织完善的

样本调查，人口普查局每月都为劳工统计局进行一次[①]。在每月含19 日的那一周里，一个由 1 500 名访问员组成的实地工作班子向约 115 000 人的一个全国性概率样本征询意见。并根据调查结果，对劳动力的规模，失业率以及其它大量的经济和人口调查的统计量(譬如，收入与受教育水平的分布)作出估计，80 年代末期，每年在这方面的花费约 3 千万美元。调查结果发布于

- 劳工评论月刊。
- 就业与收入(月度)。
- 现场人口报告(不定期)。
- 特别劳动力报告(不定期)。
- 美利坚合众国统计摘要(年度)。
- 总统经济报告(年度)。

图 1 1929 年至 1988 年失业率



来源：就业与收入，1976 年 1 月，表 A—1；1989 年 7 月，表 A—3。

本章的目的是从问题的提出开始，详细介绍现场人口调查。这将阐明和巩固前几章所介绍的思想。并亦必使其它大规模调查易于理解。这种实况研究的主要结果：

- 在实际应用中，必定使用较为复杂的概率方法抽取样本。简单随机抽样只不过是这些设计中的一个建筑版块。
- 简单随机样本的标准误差计算公式不适用于这些复杂设计，从而必须用其它方法估计标准误差。

2. 现场人口调查的设计

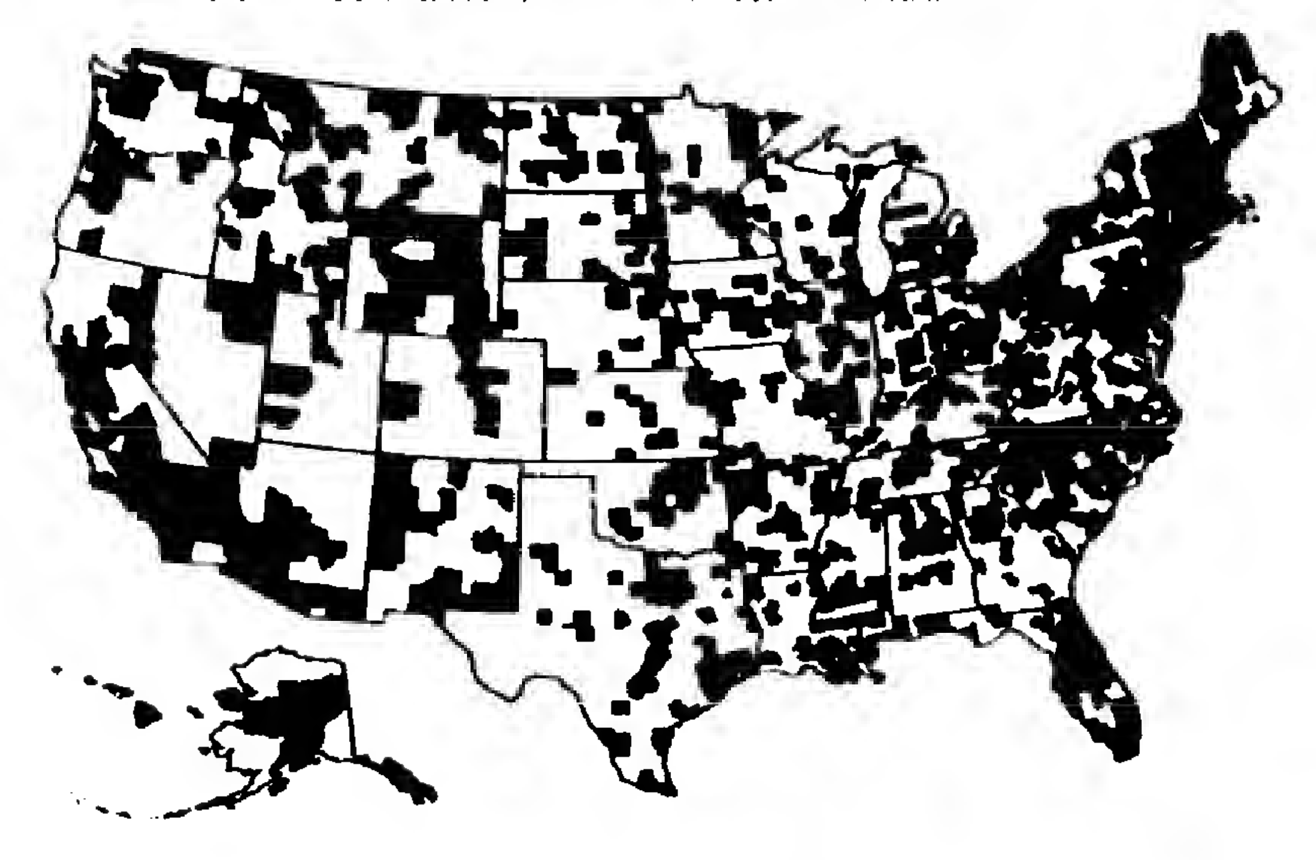
人口普查局定期重新设计现场人口调查以便适时利用新的信息和完成新的目标。80 年代初，利用 1980 年人口普查数据，曾有过一次重要的重新设计。目的是要在作出州级的各就业估计中达到较高的效率。当时全美国有 3 137 个县和独立市。作为重新设计进程的第一步，人口普查局将它们分成组群以构成 1 973 个初级抽样单元(或简记为 PSU)。每个 PSU 要么包含一个市，或一个县，或一组邻近的县[②]。这些 PSU 被划分成 715 个层，要求每个层中的 PSU 在某些人口和经济特征上(如人口变化率，或农业工人数)彼此相近。这些层不跨州。某些最大的 PSU，如纽约或洛杉矶自成一个层。

样本按二阶段选取，起步是，先从每个层中取一个 PSU，选取依概率方法进行，要求保证层中的 PSU 进入样本的机会与其人口成比例。由于有 715 个层，第一阶段产生一个 715 个 PSU 的样本。直到下一次重新设计(90 年代初)，调查的所有访问都在这 715 个 PSU 中进行，不在其它的 PSU[③]中。1980 年样本的 PSU 如图 2 所示。

每个 PSU 划分成许多基本抽样单元(或简记为 USU)，每个由约 4 个住房单元组成。第二阶段，随机取若干 USU 构成样本。最后，生活在入选的 PSU 中的被选入的 USU 中的每一位年满 16 岁或以上的人进入现场人口调查。抽样程序被设计成使得在一个州内，目标总体的每一成员有相等的机会进入样本。对整个美国来说，抽样比率约为 1 500 中取 1；但比率从最小州的 200 中取 1 变化到最大州的 2 500 中取 1[④]。

设计的目标是要求以接近相同的精度估计 50 个州和哥仑比亚特区(District of Columbia)中各自的失业人数。这等价于要求 51 个子样本的绝对容量大体相等(第 20 章的第 3 节)。因此，各州的样本容量与总体容量的比率必然有区别[⑤]。

图 2　初级抽样单元——现场人口调查(1980)。



来源:人口普查局,统计方法处。

80 年代末人口普查局的样本选择在 80 年代初就已全部产生;设计甚至为那些即将迁入还在建造中的住房的人做了准备。事实上,人口普查局不是只选择一个而是 16 个不同的样本,以便每月交替轮换样本的一部分。当一个住房单元被放进样本中后,它将被保留在样本中 4 个月,然后被换离 8 个月,而后再被送回逗留最后的 4 个月。

为什么要交替轮换样本?一个原因是,访问员可能不要很长时间就会因为频繁来访而不受欢迎。此外,作为被询问,人们的反应可能发生变化而逐渐使样本产生有偏性(称为小组偏性)的结果。譬如,有证据表明人们在第一次接受询问时比第二次更可能说他们正在找工作。那么,为什么不每个月都把样本全换掉呢?保留一部分不变可省许多钱。此外,样本中有某些交迭部分更易于估计就业和失业状态的月变化。

3. 实施调查

80 年代末的调查设计产生了约 67 000 个每月要向它们征求

意见的住房单元，其中约 11 000 个对样本来说是不可取的（从样本被派定以来，就是空的或甚至被拆毁了）。另外 3 000 左右无法利用，因为没有一个人在家，或因为在家的人不愿意合作。这样一来在调查中约留有 53 000 住房单元。这些住房单元中的年龄在 16 岁或以上的所有人都被询问有关他们上周的工作经历。在他们回答的基础上，将他们分类为：

- 受雇的（指那些上周从事任何一种有报酬工作的人，或在某家庭业务至少工作 15 小时者，或离开固定工作的休假者）；
- 或失业的（指那些上周没有受雇但可以工作且过去四周来一直在找工作的人）；
- 或劳动力之外的（蔑视亚里士多德哲学的提法，人口普查局将这定义为既非受雇亦非失业状态）。

对受雇者，询问有关他们的上班时间以及他们的工作类型。对失业者，询问有关他们的最后职业、何时何因离职以及他们正如何找工作。对那些劳动力之外的人，询问他们是否料理家务，或上学，或不能工作，或由于某种原因而不工作（在这种情形，要求他们详细说明）。1989 年 6 月公布的调查结果如表 1 中所示。

表 1 年龄在 16 岁或以上的平民非公共机构人口[6]。1989 年 6 月劳工统计局估计。

单位：百万

受雇的	118.7		
失业的	6.9		
劳动力		125.6	
劳动力之外的		60.7	
总数			186.3

来源：《就业和收入》，1989、6 月，表 A-4。

根据定义，民用劳动力包括受雇的或失业的平民。（译者注：指非军人）1989 年 6 月，总计为 118.7+6.9=125.6 百万人[7]。失业率为失业民用劳动力的百分数，故为

$$\frac{6.9}{125.6}\times 100\%\approx 5.5\%$$

这个 5.5% 是总体的全部子群组失业率的平均数。象许多平

均数一样，它隐含某些惊人的差异。这些差异将由交叉列表过程显示。失业更多地落到十几岁的少年、妇女和黑人身上，如表 2 所示。

表 2 按种族、年龄和性别的失业率。1989 年 6 月劳工统计局估计。

单位：百分比

种族	性别	16—19	年龄组 20—64	65 及以上
白人	男	14.4	3.5	2.4
白人	女	15.0	4.1	2.4
黑人	男	36.4	9.6	1.8
黑人	女	46.4	10.5	4.2

来源：《就业与收入》，1989、6 月，表 A-4。

如图 1 所示，综合失业率变化相当大。但是表 2 中的失业率型式却相当稳定。例如，在 1961—1989 这期间，黑人的失业率大体上为白人失业率的两倍。

失业人数以一种比表 2 所示的更为完善的分类方式公布，它包括：婚姻状况，最后职业类型，失业原因（被解雇或辞职），以及"职业寻求方法"。劳工统计局从样本中的约 115 000 人开始。直到辞去他们的职务并通过查阅招聘广告谋职的年龄在 45—54 的白人，不可能有许多状态剩下了（图 3）。



总之，当出现一个大的样本被交叉列表时，在某些分类中可能只有非常小的子样本。从而有关相应子总体的推断会变得颇为靠不住。但是，譬如说，甚至假设每个估计以概率95%在其真值的1%之内。对一千个估计(差不多是《就业与收入》中的数目)来说，它们中的少数几个出现比1%大约多的偏离并不足为奇。劳工统计局取了一个大样本，因为它必须对许多子总体作出许多估计，并且它希望有理由相信它们都是相当精确的。事实上，当子样本降到低于约50的容量阀值时，对这些状态劳工统计局将不做估计。图3中只有少数几个这种单元：它们都标有(′)。(见下页表)

4. 样本加权

假设某月劳工统计局的115 000人的样本里有4 536人失业。劳工统计局是按1 500中抽1人(抽自16岁或以上的平民非社会事业机构人口)。因此，很自然会想到样本中的每个人代表国内的1 500人。从而估计人口总体中总的失业人数的方法相应为用因子1 500对样本失业人数4 536加权：

$$1\ 500\times4\ 536=6\ 804\ 000$$

但是，劳工统计局并不这样简单行事。不是样本中的每个人得到相同的权。劳工统计局将样本(按年龄，性别，种族和居住区域)分成群组，并对各群组分别加权。

对这种复杂化有一个充分的理由：权重有助于控制机会变差的影响。譬如，相对于在总人口中所占的份额，假设样本中有太多的16—19岁的白种男人，若这群组的失业率偏高，则会导致样本的综合失业率过高。为了补偿，样本中过份体现的群组应按比例赋予较小的权，使样本回复到与总体一致。另一方面，若一个群组体现不足时，所加的权应略为增大。按这种方式调整权重有助于纠正机会变差所引起的不平衡，并减少抽样误差[⑧]。

表 3 1989，6 月，《就业与收入》第 28 页，表 A-19。

A-19，按性别、年龄、种族和所用求职方法的失业求职者

性别、年龄和种族	1989 年 6 月								平均使用方法个数
	千人		采用这些方法占求职总人数的百分比						
	失业总人数	求职总人数	政府职业介绍所	私人职业介绍所	与雇主直接联系	登广告或对广告应聘	朋友或亲戚	其它	
总人数，16 岁及以上	6 850	5 991	19.3	7.2	71.0	37.8	19.5	5.6	1.60
16 到 19 岁	1 687	1 627	12.0	3.1	76.3	26.8	17.1	5.0	1.40
20 到 24 岁	1 372	1 232	19.8	7.1	72.6	38.6	17.5	4.4	1.60
25 到 34 岁	1 805	1 517	22.3	7.7	71.1	42.8	20.5	4.8	1.69
35 到 44 岁	917	825	26.1	8.8	65.2	47.2	25.9	6.9	1.80
45 到 54 岁	549	427	24.8	12.4	65.6	40.7	16.9	8.7	1.69
55 到 64 岁	382	295	15.6	13.2	60.7	42.0	19.7	10.2	1.61
65 岁及以上	83	67	(')	(')	(')	(')	(')	(')	(')
男人，16 岁及以上	3 484	2 984	20.1	6.7	70.9	37.0	22.4	7.0	1.64
16 到 19 岁	848	821	14.0	3.4	73.6	26.8	18.6	6.5	1.43
20 到 24 岁	702	618	20.4	6.1	72.5	38.5	21.2	4.7	1.63
25 到 34 岁	912	718	22.7	7.4	70.8	44.0	23.1	6.1	1.74
35 到 44 岁	496	413	26.4	8.2	71.9	42.9	32.2	6.8	1.88
45 到 54 岁	274	212	26.4	11.3	65.1	37.3	24.1	13.7	1.78
55 到 64 岁	204	162	15.4	9.3	58.0	42.6	17.3	16.0	1.59
65 岁及以上	48	39	(')	(')	(')	(')	(')	(')	(')

续表

性别、年龄和种族	1989年6月								平均使用方法个数
	千人		采用这些方法占求职总人数的百分比						
	失业总人数	求职总人数	政府职业介绍所	私人职业介绍所	与雇主直接联系	登广告或对广告应聘	朋友或亲戚	其它	
妇女,16岁及以上	3 365	3 007	18.6	7.7	71.1	38.5	16.6	4.2	1.57
16到19岁	839	806	10.2	2.6	79.2	26.8	15.8	3.6	1.38
20到24岁	670	614	19.2	8.1	72.8	38.6	14.0	4.1	1.57
25到34岁	893	799	22.0	8.0	71.3	41.7	18.1	3.6	1.65
35到44岁	474	411	25.8	9.5	58.4	51.6	19.5	7.1	1.72
45到54岁	275	215	23.7	13.5	66.0	44.7	10.2	4.2	1.62
55到64岁	179	133	15.8	18.8	63.9	41.4	22.6	3.0	1.65
65岁及以上	35	28	(')	(')	(')	(')	(')	(')	(')
白人,16岁及以上	4 893	4 166	18.7	7.1	71.6	40.0	19.6	6.1	1.63
男人	2 545	2 114	19.8	6.2	72.2	38.8	22.3	7.1	1.66
妇女	2 348	2 051	17.5	8.0	71.0	41.4	16.7	5.0	1.60
黑人,16岁及以上	1 728	1 624	20.0	7.5	71.2	31.1	19.1	4.2	1.53
男人	819	767	19.8	7.6	70.9	30.8	22.6	6.0	1.58
妇女	909	858	20.0	7.3	71.6	31.4	16.0	2.7	1.49

5. 标准误差

在估计失业率中，要考虑精确性。譬如，可以用报道失业率为7%±(1%的0.1)来勾划经济的确切写照。不管怎么说，7%±2%把从繁荣到失败的一切都包括进去了。因此，了解估计确实有多好是很重要的。由于劳工统计局没采用简单随机样本，因此，我们到目前为止所导出的方法不适用。

特别在抽样程序的第二阶段，劳工统计局选取某些基本抽样单元(USU)：一个USU是约四个毗连住房单元的整群。凡生活在这些USU中每一个年龄在16岁和以上的人都进入样本(第2节)。因此，一个整群为全是或全不是：要么整群中的每个人都进入样本，要么谁也不进。于是，生活在同一整群中的人们在许多方面趋于互相类同；从家庭背景，学历，就业状况等方面，有关每一个人的信息述说了有关所有其它人的一些东西。

相比之下，对于简单随机样本，若街区的某人进入样本，其邻居仅有很小的机会进入样本。因此，每选一个新的人进入简单随机样本都提供另外的信息，独立于前面已入选的人。对于整群抽样样本，存在许多多余信息。因此，劳工统计局的115 000人的整群样本所含的信息比同容量的简单随机样本的少。

> 整群样本比同容量的简单随机样本的信息少。因此简单随机样本有关标准误差的公式不适用。

整群式降低了劳工统计局估计的精确性。另一方面，加权又改进了精确性。总而言之，计算劳工统计局的估计的SE是件棘手的事。

事实上对于整群样本，标准误差本身可用半样本法由数据非常接近地估计出。尽管细节相当复杂且需要大量的计算，但想法简单。如果劳工统计局想知道现场人口调查有多精确，要做的是按照完全相同的程序再独立做一次调查。两次调查的差异会给你一些每组结果有多可靠的概念。

没有人会仅仅为了看一下它有多可靠，而认真地建议每年再多花费3千万美元去重复做一次现场人口调查。但是，劳工统计局可以通过将现场人口调查样本剖分成相互独立，具有相同机会习性的两部分而获得几乎相同的效果。（故取名“半样本法”。）譬如，假设调查的一部分估计民用劳动力为125.5百万，另一部分估计为125.7百万，差异是由机会误差产生的。民用劳动力的共同估计为

$$\frac{125.5+125.7}{2}=125.6\text{ 百万}$$

两个各别估计偏离它们的平均数0.1百万，标准误差就用这个0.1百万的差来估计。

当然，仅基于一次剖分来估计的标准误差可能并不太可靠。但是有许多不同的剖分样本的方式。劳工统计局考虑了一些剖分样本的方式并利用取均方根将相应的标准误差结合起来，这就完成了半样本法的轮廓[9]。对1989年6月标准误差的一些估计列于表3中。

表3 估计的标准误差，1989年6月。

	估计	标准误差
民用劳动力	125.6百万	275 000
就业人数	118.7百万	293 000
失业人数	6.9百万	136 000
失业率	5.5%	1%的0.1

来源：就业与收入，1989年7月，表B和C，第164页。

表3中估计的标准误差与同容量、同成分的简单随机样本的标准误差相比较情况如何呢？计算表明，对估计劳动力的规模，劳工统计局的标准误差比简单随机样本的要小约5%：加权起了好作用。但是，对于估计失业人数，劳工统计局的样本比简单随机样本约差30%：整群方法造成危害[10]。

那么，为什么劳工统计局不使用简单随机样本呢？原因之一，没有一份全美国16岁及以上的人的名单和他们的现地址。即使有

这样一份名单，从中抽取一个简单随机样本将产生稀稀拉拉地散布在全国各地的人选，访问他们的花费会非常大。采用劳工统计局的办法，样本必定成群地在相对小的适当限定的区域内产生，因此访问费用相当容易安排：80 年代末，每户约 20 美元。劳工统计局精心的样本设计出现令人惊奇的费用效率。

劳工统计局的设计和简单随机样本之间的相互比较指出了真正的争论焦点。为了适当地计算标准误差，你需要样本数据之外更多的东西：你需要知道样本是如何选取的。对于简单随机样本，存在一个 SE。对于整群样本，存在另一个。（这些争论点前面也出现过，在盖洛普民意测验的有关章节中：第 21 章的第 4、5 节。）

“方便样本”是不用概率方法选取的样本。（例如，某讲师的一年级心理学班级。）有些人把简单随机样本的公式用于方便样本。那可能是个真正的大错误。对于方便样本，是很难定义机会的；参数和标准误差亦如此。

> 标准误差的公式必须注意考虑用以抽取样本的概率方法的细节。对于方便样本，标准误差一般无意义。

习题 A

1.（假设的）某市卫生部门取一个 100 户的简单随机样本。其中的 80 户其所有居民都接种了防小儿麻痹症的疫苗。因此，该部门估计该市 80%±4% 左右的住户的所有居民都接种过防小儿麻痹症的疫苗，4% 是正确的 SE 吗？

2.（续题 1.）该部门访问了样本家庭中每位 25 岁及以上的人。他们发现共有 144 位这类人，其中 29 位有大学文凭。他们估计该市 25 岁及以上的人口中 20% 有大学程度。

(a) 这是一个整群样本吗？

(b) 你能根据给定的信息估计百分数的 SE 吗？

3. (a) 某心理学讲师将他班上的全体学生取为他的样本，这是一个整群样本

吗？

(b)某社会学家询问了某日穿过购物中心的头 100 人。这是一个整群样本吗？

4. 某月，现场人口调查样本共计有 100 000 人，其中 60 000 人受雇，5 000 人失业。正确还是错误，并解释之：

(a)样本的 65%是劳动力。

(b)劳工统计局将估计人口的 65%属于劳动力。

5. 设将现场人口调查样本剖分成两个独立的一半。根据其中一半，估计受雇人数为 111.5 百万；根据另一半，估计值为 111.3 百万。试将这两个估计结合起来，并附带估计结果的标准误差。

6. 在选举年里，劳工统计局用现场人口调查样本对投票情况写一份特别报告。在 1984 年，样本中所有法定投票年龄的人中的约 68%说他们投了票；但是法定投票年龄的总人口中只有 55%实际投票了[11]。这差异能解释成机会误差吗？若不能，它还能如何解释？(你可以设劳工统计局的样本等价于 75 000 人的简单随机样本。)

7. 在表 2 中，哪一个估计更可信：20—64 岁的白人男性的，或 20—64 岁的黑人男性的？简要解释之。

这些习题的答案在第 714 页。

6. 数据的质量

现场人口调查所收集的数据有很高的质量。譬如，他们的数据就被认为比人口普查的数据精确。在任何大规模现场操作中，错误是不可避免的；由于现场人口普查调查在一个远小于人口普查的规模上操作，因此它能提供较好的质量控制。关键在于仔细挑选、培训和指导现场工作人员。譬如访问员，在他们开始上岗工作之前，给他们约四天时间的调查程序培训，在他们上岗之后，每月给他们若干小时的培训。他们的主管人每年至少要对他们的工作考查一次。另外，约 3%的月样本(由单独的概率抽样程序选取)由主管人员重新访问。所有不相符之处都与访问员讨论。访问员的报告全加以校正，即查核不完全和不一致的记载。对大部分项目来说，差错率是低的；且对每项差错都与造成它的人员复核一下。

7. 偏性

偏性比机会误差更不易觉察到其危害,特别是如果它多多少少均匀地在整个样本起作用。由半样本法——或任何其它方法——计算的 SE 将不能探查出这类偏性。度量偏性,即使是粗糙地,也是件困难的事,且必须包括超出样本数据之外的内容。

> 当偏性多多少少均匀地散布在整个样本时,它不可能仅通过查看数据检出。

劳工统计局对现场人口调查中的偏性做了不同寻常的仔细研究。总的说来,这些偏性看来是较小的,尽管它们的确切大小并不知道。首先,现场人口调查设计是基于人口普查数据(第 2 节),而人口普查数据缺失总人口的一个小的百分数。这百分数不容易定下来[12]。即使劳工统计局知道这一点,他们很难调整失业的估计人数(譬如说)以补偿计数不足部分,因为人口普查数据缺失的人可能与他们查找到的人有很大的区别。这种困难也会在其它地方突然冒出。现场人口调查约缺失人口普查数据所计数人口的 7%。在某种程度上,权可将缺失的人送回样本中(见第 4 节)。但不回答的偏性仍可能存在:现场人口调查所缺失的人可能略有别于他们查找到的人,权把他们打扮成相同的。

其次,"受雇"与"失业"之间的区分边界有点模糊。例如,打零工的人会象有专职工作的人一样被分类为受雇的,但他们确实是部分失业的。另一点:想工作但已放弃找工作的人被分类为劳动力之外,尽管他们可能应该分类为失业者。劳工统计局的失业标准,即尚无工作、可以干工作、且正在找工作,必定是主观的。在实际应用中,它有点含糊。重新访问措施的结论(第 6 节)提出失业人数比劳工统计局的估计约高几十万人左右。在这种情况,偏性比抽样误差大[13]。有一个劳工统计局不存在的问题:家庭偏性(第 19 章的第 4 节)。理由是样本包括了选取家庭中的全部 16 岁及以上的人,而不只是访问员在家里找到的一个人。

8. 复习题

复习题同样可能涉及前几章。

1. 某月，现场人口调查样本中有 100 000 人，其中 60 000 受雇，4 000 失业。

   (a)正确还是错误，并解释之：劳工统计局可能估计失业人口的百分数为

$$\frac{4\ 000}{60\ 000+4\ 000}\times 100\%\approx 6.2\%$$

   (b)其余 36 000 人出现了什么情况?

2. 某月，现场人口调查样本中有 100 000 人，劳工统计局估计失业率为 6.0%。正确还是错误，并解释之：这百分数的标准误差应估计如下——

$$\text{人数的 SE}=\sqrt{100\ 000}\times\sqrt{0.06\times 0.94}\approx 75$$

$$\text{百分数的 SE}=\frac{75}{100\ 000}\times 100\%\approx 1\%\text{ 的 } 0.08$$

3. 某月，现场人口调查样本被剖分成两个独立相同的样本。利用其中一个，失业人数估计为 7.1 百万。另一个产生一个 6.9 百万的估计值。利用这个信息，试估计失业人数，并对该估计附加估计标准误差。

4. 利用习题 3 中的数据，对估计中的偏性能说些什么?

5. 某超级市场连锁店在每年年底必须对存货作价，一般在样本基础上进行。该店中出售的所有类型的商品品种有一份总的清单。于是，查帐人从所有货架中取一个商品样本，并找出样本商品的价格和存货数量。为了抽取样本，查帐人从数 1 到 100 中随机取一个数。假设这个数是 17。查帐人制取清单上第 17，第 117，第 217，……个商品品种为样本。若该随机数是 68，他们则取第 68，第 168，第 268……个品种。等等。

   (a)这是一种概率方法吗?

   (b)这是一个简单随机样本吗?

回答是或不是，并解释之。

6. (假设的)原告律师协会估计其成员的10%赞成无过失自动保险。这个估计是基于出席某次全国性会议的成员所填写的2 500份问卷。正确还是错误，并解释：这个估计的SE为1%的0.6，因为

$$\sqrt{2\ 500}\times\sqrt{0.1\times 0.9}=15,\ \frac{15}{2\ 500}\times 100\%=1\%\text{的 }0.6$$

7. 某调查机构取一个简单随机样本来估计某镇上民主党人的百分数。他们发现这个百分数的一个95%置信区间为49%±6%。于是，这百分数的一个99.7%置信区间为：

   (i)不可能求出，因为样本容量未知，

   (ii)$49\%\pm\frac{100-95}{100-99.7}\times 6\%$，

   (iii)$49\%\pm 2\times 6\%$，

   (iv)$49\%\pm\frac{99.7}{95}\times 6\%$，

   (v)$49\%\pm\frac{3}{2}\times 6\%$

   取一种选择，并解释之。

8. 根据《旧金山记事报》中L. M. Boyd的专栏"The Grab Bag(百宝囊)"："平均数律说如果你掷一对骰子100次，掷出的点数加起来约为683。"这是正确的吗？回答是或不是，并解释之。

9. 作为喝酒问题研究的一部分，通过访问获得了一个酗酒者态度的样本[15]。患者是随机分派给访问员的。某些访问员是绝对戒酒者，另一些则喝酒。你能期望这两组访问员会获得相似的结论吗？回答是或不是，并给出理由。

10. 简单随机样本是随机____放回地抽取。选择：有；不。

11. 一盒装有250张票。两个人想估计盒中1的百分数。他们同意用从盒中随机抽取的100张票中1的百分数。A想放回地抽取，B想不放回地抽取。哪一种方法的估计较精确？或它们有什么区别吗？

12. “往空中抛一百枚分币并记录它们落下时出现头像的枚数。重复抛几千遍并绘制所得数的直方图。你将获得一张近似于正态曲线的直方图，且你抛百枚分币的次数愈多你的直方图愈接近于曲线。”[16]若你继续不停地抛这组百枚分币，你的直方图会愈来愈接近于正态曲线吗？或它将收敛到抛一枚硬币 100 次中头像数的概率直方图吗？取一种选择，并简短地解释之。

9. 小结

1. 美国的失业率由月抽样本调查估计，称为现场人口调查。

2. 这项调查使用一个约 115 000 人的全国性概率样本。其设计比简单随机样本复杂得多。

3. 这项调查对样本重新加权使它与人口普查数据在年龄、性别、种族、居住的州，以及其它影响就业状态的某些特征方面一致。

4. 当样本依概率方法抽取时，它不仅可用来估计参数，而且也可以用来计算估计中机会误差的可能大小。

5. 整群样本的标准误差可由半样本法获得：将样本剖分成两半并察看它们有多相符。

6. 标准误差的公式必须注意考虑用以抽取样本的概率方法的细节。应用于简单随机样本的公式一般会偏低估计现场人口调查中所用的那种整群样本的标准误差。

7. 对于方便样本，标准误差一般无意义。

8. 当偏性多多少少均匀地散布于整个样本时，不可能通过查看样本数据来检测它。标准误差对这类偏性忽略不计。

9. 现场人口调查，象所有调查一样，受一些小的偏性影响。失业率估计中的偏性被认为大于标准误差。

23

平均数的精度

牧场是为牛用的

——L. B. J.

1. 引言

本章的目的是估计由简单随机样本算出的平均数的精度。本节论述一个初步的问题:从盒子中摸出的数的平均数中存在多大的机会变异?

例如,取盒子

1	2	3	4	5	6	7

编测计算机程序以从这盒中随机放回地抽取 25 次:

2 4 3 2 5　7 5 6 4 5　4 4 1 2 4　4 6 4 7 2　7 2 5 7 3

这些数之和为 105,因此它们的平均数为 105/25=4.2。计算机重复一次这种试验,产生不同的结果:

5 1 4 3 4　5 2 1 7 7　1 2 3 2 4　7 1 6 5 3　6 6 3 3 4

现在和为 95,故平均数为 95/25=3.8。抽得数之和受机会变异支配,因此平均数亦然。新的问题是算出抽得数的平均数的期望值和标准误差。将通过例子来讲解方法。

例 1. 设从盒子

1	2	3	4	5	6	7

随机放回地抽取 25 次。抽得的平均数将约为________，±______左右。

解. 盒子的平均数为 4，故抽得的平均数将约为 4。允许的误差数是 SE。为了求出平均数的 SE，我们回到和上来，和的期望值为

$$\text{抽取次数} \times \text{盒子的平均数} = 25 \times 4 = 100$$

而和的 SE 为

$$\sqrt{\text{抽取次数}} \times \text{盒子的 SD} = \sqrt{25} \times 2 = 10$$

和将约为 100，±10 左右。

这对抽得的平均数说明了什么？若该和高出期望值一个 SE，即 100+10，则 25 个抽得的平均数为

$$\frac{100+10}{25} = \frac{100}{25} + \frac{10}{25} = 4 + 0.4$$

另一方面，若该和低于期望值一个 SE，即 100－10，则平均数为

$$\frac{100-10}{25} = \frac{100}{25} - \frac{10}{25} = 4 - 0.4$$

故抽得的平均数将约为 4，±0.4 左右。4 为抽得的平均数的期望值。0.4 为标准误差。解毕。

要求出平均数的 SE，需追溯求得和的 SE，再除以抽取次数。平均数的 SE 述说了抽得的平均数与盒子的平均数可能相差多远。

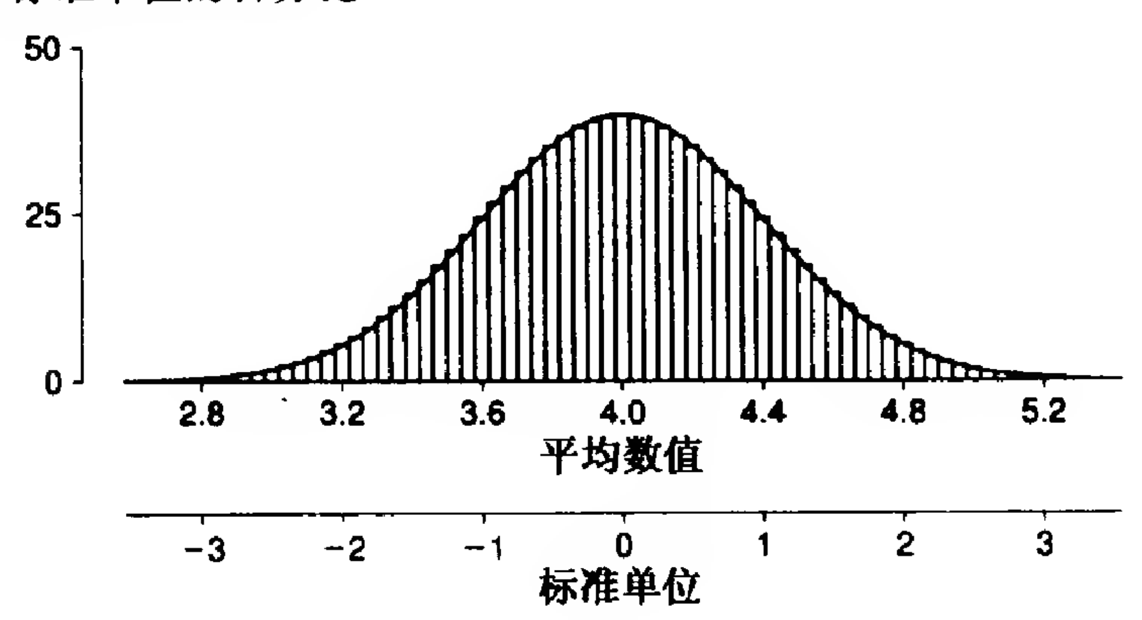
> 当从某盒中作随机抽取时，抽得的平均数的期望值等于盒子的平均数。抽得的平均数的 SE 等于
>
> $$\frac{\text{它们的和的 SE}}{\text{抽取次数}}$$

若抽取次数足够大，正态曲线可用来估计平均数的机会。图 1 展示了从盒子

[1][2][3][4][5][6][7]

中抽取 25 个数的平均数的概率直方图。直方图遵循正态曲线，因此直方图下的面积可由正态曲线下的面积近似。

图 1　从盒子[1][2][3][4][5][6][7]中抽取 25 个数的平均数的概率直方图。直方图遵循正态曲线(绘出供比较)。纵向尺度是按每标准单位的百分比。



为什么平均数的概率直方图看起来象正态曲线？这由第 18 章中的数学而得。有了足够多的抽取次数，和的概率直方图将接近正态曲线。抽得的平均数等于它们的和除以抽取次数。这一除法运算只是尺度上的改变，并在标准单位中消除[1]。

当从盒子中随机抽取时，抽得的平均数的概率直方图将遵循正态曲线，即使盒子里的数值并不如此。直方图必须换算成标准单位，且抽取的次数必须充分大

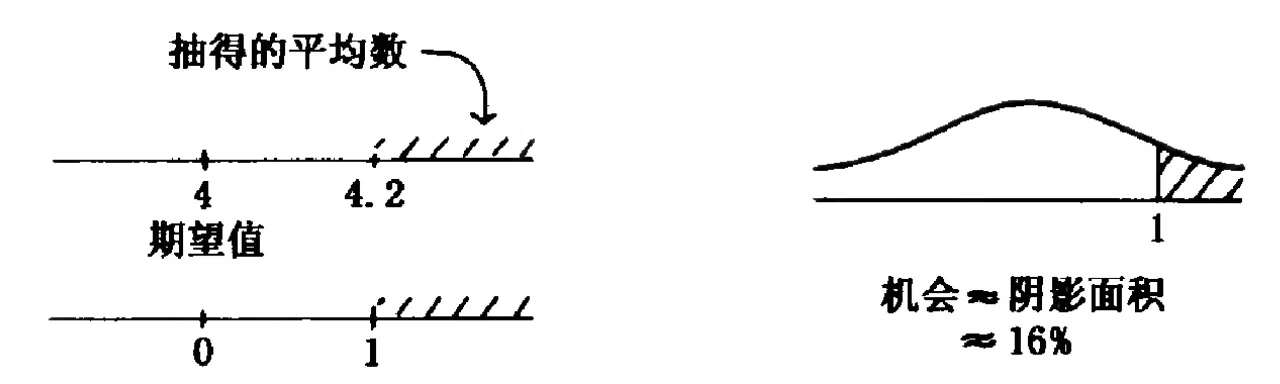
例 2. 设将从例 1 中的盒子里随机放回地抽取 100 次。

(a)抽得的平均数将约为________，±________左右。

(b)试估计抽得的平均数将大于 4.2 的机会。

解　如在例 1 中，抽得数之和将约为 $100\times4=400$，加或减去 $\sqrt{100}\times2=\pm20$ 左右。抽得的平均数将约为 $400/100=4$，加或减去 $20/100=0.2$ 左右。100 次抽得的平均数的 SE 为 0.2。

(b)可用正态近似处理。



机会约为 16%。解毕。

在例 1 和 2 中，若抽取次数按因子 4 增长，从 25 增长到 100，抽得的平均数的 SE 按因子$\sqrt{4}=2$下降，从 0.4 下降到 0.2。一般如此。

当从一只装有票子的盒子中随机放回地抽取时，若抽取次数乘以某一因子(譬如 4)，则抽得的平均数的 SE 除以这因子的平方根($\sqrt{4}=2$)。

当抽取次数增加时，和的 SE 变大——而平均数的 SE 变小。理由如下。和的 SE 增大，但仅按抽取次数的平方根增长。因此，尽管和的 SE 按绝对意义增大了，但与抽取次数相比较它变小了。于是，用抽取次数相除使平均数的 SE 下降。记住这两种 SE 之间的差别。

当抽取是不放回时，抽得的平均数的确切 SE 可用修正因子获得(第 20 章，第 3 节)——

不放回的 SE=(修正因子)×(放回的 SE)。

通常，抽取次数与盒中的票子总数相比较要小得多，从而修正因子将很接近于 1，故可忽略不计。

习题 A

1. 从某盒中随机放回地抽取 100 次。

(a)若抽得数之和为 110，它们的平均数是什么?

(b)若抽得数的平均数为 0.9，它们的和是什么?

2. 试填下表，对于从盒子 [0] [2] [3] [4] [6] 中随机放回抽得的数值

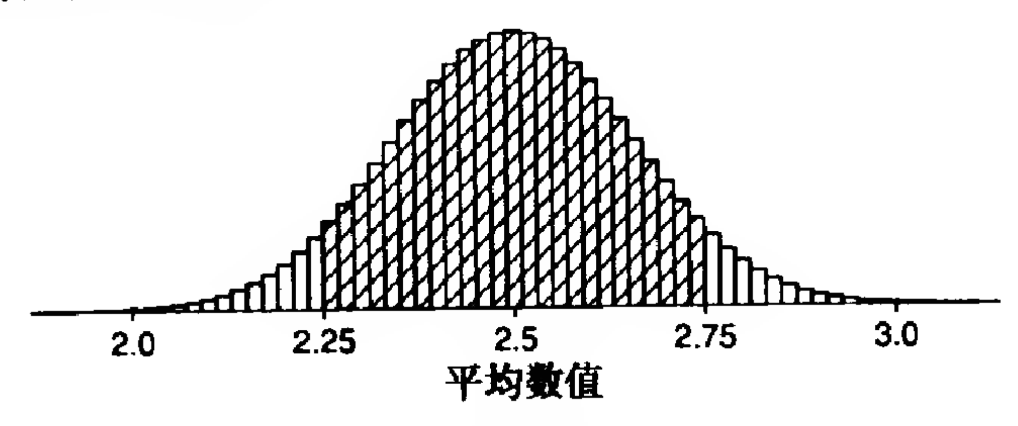
抽取次数	抽得数之和的 SE	抽得数的平均数的 SE
25		
100		
400		

3. 某盒有 10 000 张票。这些票上的数平均为 50,SD 为 20。

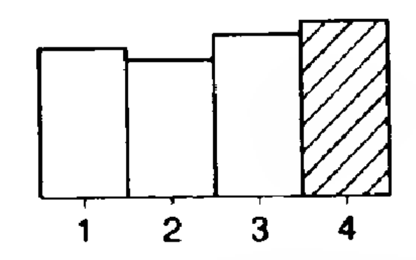
(a)随机放回地抽取 100 张票。这些抽得数的平均数将约为________,±________左右。

(b)若抽取是不放回的，又如何?

4. 下图是从盒子 [1] [2] [3] [4] 中抽取的 50 个数的平均数的概率直方图。阴影部分的面积表示什么?



5. 下图是从题 4 中的盒子里抽取 50 次产生的数据的直方图。阴影部分面积表示什么?



6. 一盒票的平均数为 75,SD 为 10。从这盒中随机放回地抽取 100 次。

(a)试求抽得的平均数将在 65 到 85 范围中的机会。(近似地)。

(b)再对范围 74 到 76 做一遍。

7. 设从某票盒中随机放回地抽取 100 次。盒中的平均数为 200。抽得的平均数的 SE 经计算为 10。正确还是错误:

(a)盒中约 68%的票在范围 190 到 210 中。

(b)对 100 次抽得的平均数，有约 68%的机会在范围 190 到 210 中。

8. 有三只装有标上数值的票的盒子，每只盒子里数的平均数都为 200。但是，A 盒的 SD 为 10，B 盒的 SD 为 20，C 盒的 SD 为 40。现在，

- 从 A 盒中抽取 100 次
- 从 B 盒中抽取 200 次
- 从 C 盒中抽取 400 次

(抽取是放回的。)计算了各组抽得数的平均数，打乱次序后为，

203.6　　198.1　　200.4

(a)哪一个平均数来自哪一只盒?

(b)它可能会是别的吗?

简短地解释之。

这些习题的答案在第 714—715 页。

技术性注释。当从某盒子中随机放回地抽取时，抽得数之和的 SE 为

$$\sqrt{\text{抽取次数}} \times \text{盒子的 SD}。$$

故抽得的平均数的 SE 为

$$(\sqrt{\text{抽取次数}} \times \text{盒子的 SD})/\text{抽取次数}。$$

它可简化为(盒子的 SD)/$\sqrt{\text{抽取次数}}$，在大多数书中表为 $\sigma/\sqrt{n}$，这里 σ 为 SD，n 为抽取次数。希腊字母 σ 念成“sigma”。

2. 样本平均数

在第一节中，盒中的数是已知的，问题是对抽得的平均数作出某种论述。本节从相反——而更实用——的方向推理：从某成份未知的盒子中抽取一个随机样本，问题是对盒子的平均数作出估计。自然抽得数的平均数可用作估计值，且样本平均数的 SE 可以结合正态曲线用来评估估计的精度(第 21 章中对百分数使用了相同的方法。)

将用例子介绍方法。同时，有两个问题需要回答：

- 样本的 SD 与样本平均数的 SE 之间有什么区别?
- 为什么在描绘置信水平时使用正态曲线是合理的?

先看例子。设某市行政官想知道生活在他的城镇的 25 000 户

家庭的平均收入。他雇一个调查机构抽取一个 900 户家庭的简单随机样本。900 户样本家庭的收入的平均数为 32 400 美元，SD 为 18 000 美元。在这基础上，全体 25 000 户家庭的平均收入估计为 32 400 美元。当然，这由于机会误差而有所偏离。问题是对这估计加或减一个数：

32 400 美元±________美元?

这需要用到 SE，为此需要一个盒子模型。对镇上的每户家庭，盒子中都应有一张票与之对应，票上注明该家庭的收入。数据就是从盒中抽出的 900 个数。样本家庭的平均收入等于抽得的平均数。而抽得的平均数的 SE 可以用第 1 节的方法求出。第一步，求出抽得数之和的 SE。由于 900 是 25 000 的一个很小的部分，因此在放回与不放回抽样之间并无实质性差异。故和的 SE 为

$\sqrt{900}\times$ 盒子的 SD。

当然，调查机构并不知道盒子的 SD，但他们可以用样本的 SD 估计它。这是第 21 章第 1 节中所讨论的自助方法的又一例子。

对于简单随机样本，样本的 SD 可用来估计盒子的 SD。当样本容量足够大时，估计是好的。

在本例中，盒子的 SD 可由样本的 SD 估计，为 18 000 美元。故和的 SE 估计为

$\sqrt{900}\times 18\ 000$ 美元 $=540\ 000$ 美元。

为了求出平均数的 SE，我们将它除以样本中的家庭户数：540 000 美元/900＝600 美元。这是答案。抽得的平均数偏离盒子的平均数大约为 600 美元。故镇上的全体 25 000 户家庭的收入的平均数可估计为

32 400 美元±600 美元。

记住对这 600 美元所作的注释：它是估计的误差的幅度。例子完毕。

第 21 章的第 2 节讨论了百分数(定性数据)的置信区间。类似

想法可以用来求盒子的平均数(定量数据)的置信区间。譬如,镇上全体 25 000 户家庭收入的平均数的近似 95%置信区间可以通过从样本平均数向两边各移出 2 个 SE 求得:

32 400 美元±2×600 美元=31 200 美元至 33 600 美元。

(“样本平均数”是样本中数的平均数据的统计速记。)

计算中出现两个不同的数:样本的 SD 为 18 000 美元和样本平均数的 SE 为 600 美元。这两个数起不同的作用。

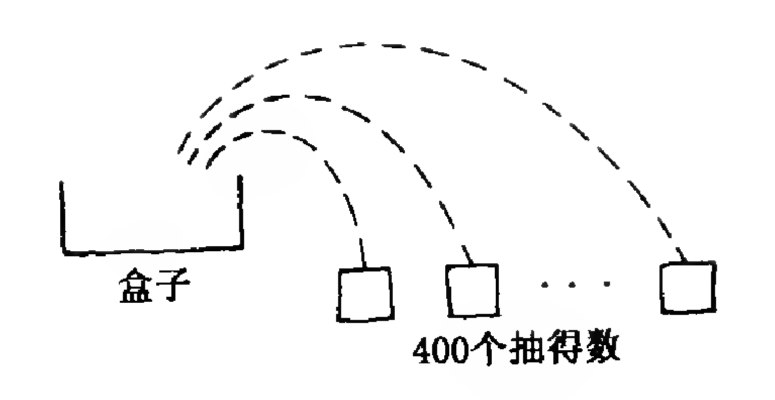
- SD 指出了一个家庭的收入离平均数——一个典型家庭的收入有多远。
- SE 指出了样本平均数离总体平均数一个典型的样本有多远。

将 SD 和 SE 搞混的人会认为不知为什么镇上 95%的家庭其收入在 32 400 美元±1 200 美元的范围内。这是荒谬的:收入的 SD 约为 18 000 美元。置信区间是指非常不同的东西:对所有抽样的约 95%来说,若你从样本平均数向两边各移 2 个 SE,你所得到的置信区间将覆盖全镇的平均数;而你的置信区间里没有另外的 5%。“置信”一词意在提醒你,机会存在于抽样程序中;盒子的平均数不会变动。(这些问题在前面第 21 章的第 3 节中讨论过。)

例 3. 作为民意调查的一部分,在某镇抽取了一个 400 名 25 岁和以上的人的简单随机样本。样本中诸人的平均教育水平(完成正规教育的年数)为 11.6 年,SD 为 4.1 年。试求镇上全体 25 岁和以上的人的平均教育水平的近似 95%置信区间。

解. 首先,要有个盒子模型:对镇上每位 25 岁和以上的人,盒中都应有一张票与之对应,票上注明那个人完成正规学校教育的年数。从盒中随机抽取 400 张。数据就是抽得数,样本平均数为抽得数的平均数。模型建立完毕。

要用到平均数的 SE。抽得数之和的 SE 为$\sqrt{400}\times$盒子的 SD。盒子的 SD 是未知的,但可用样本的 SD 估计,为 4.1 年。故抽得数之和的 SE 估计为$\sqrt{400}\times 4.1=82$ 年。于是平均数的 SE 为



82/400≈0.2 年。样本中诸人的平均教育水平将偏离该镇的平均数 0.2 年左右。镇上全体 25 岁和以上的人的平均教育水平的近似 95%置信区间为

$$11.6 \pm 0.4 \text{ 年}$$

这是答案。

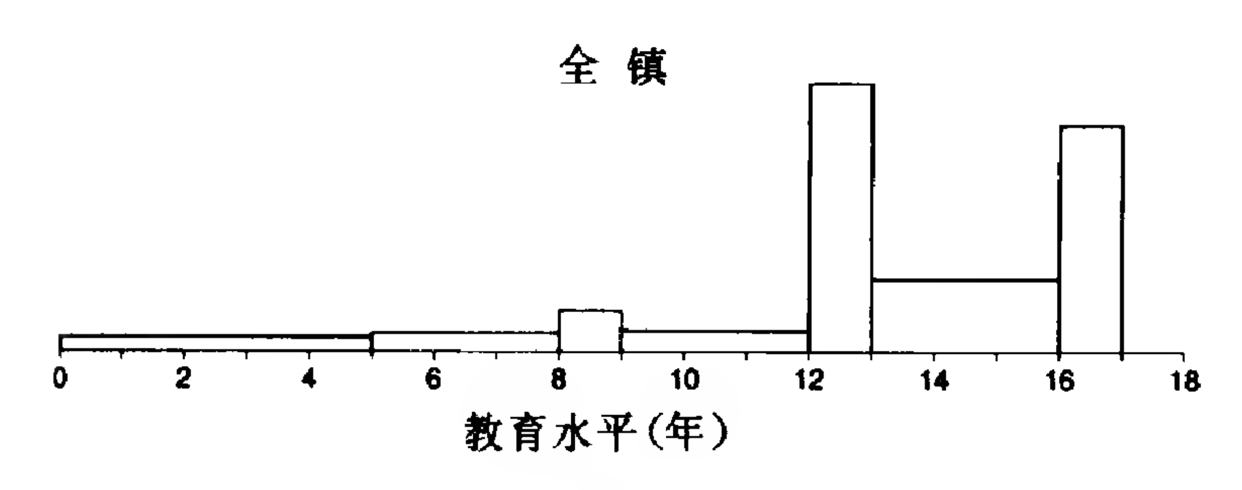
在这一点上，很自然会反对说教育水平的直方图(见 43 页)看起来一点不象正态曲线。为什么用正态曲线来描述置信水平是合理的呢？原因：该曲线不是用来近似数据的直方图的，而它是用来近似样本平均数的概率直方图的。这两种直方图之间的区别是很关键的。

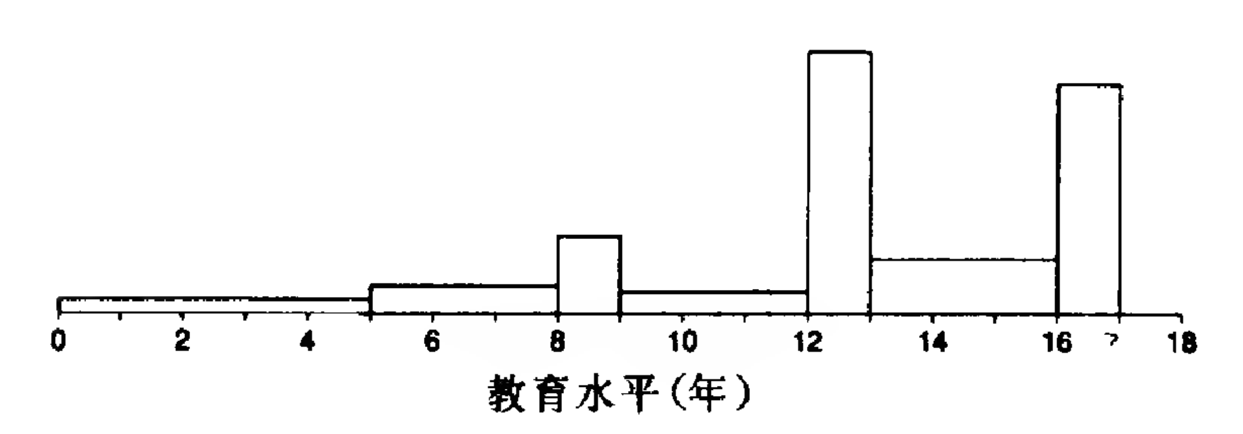
盒子模型有助于集中反映问题。别忘了对镇上每个 25 岁和以上的人，盒中都有一张票与之对应，票上注明他或她的教育水平。盒中内容的直方图在图 2 的上部给出。这张直方图描绘了镇上全体 25 岁和以上的人的教育水平。它的形状一点也不象正态曲线。

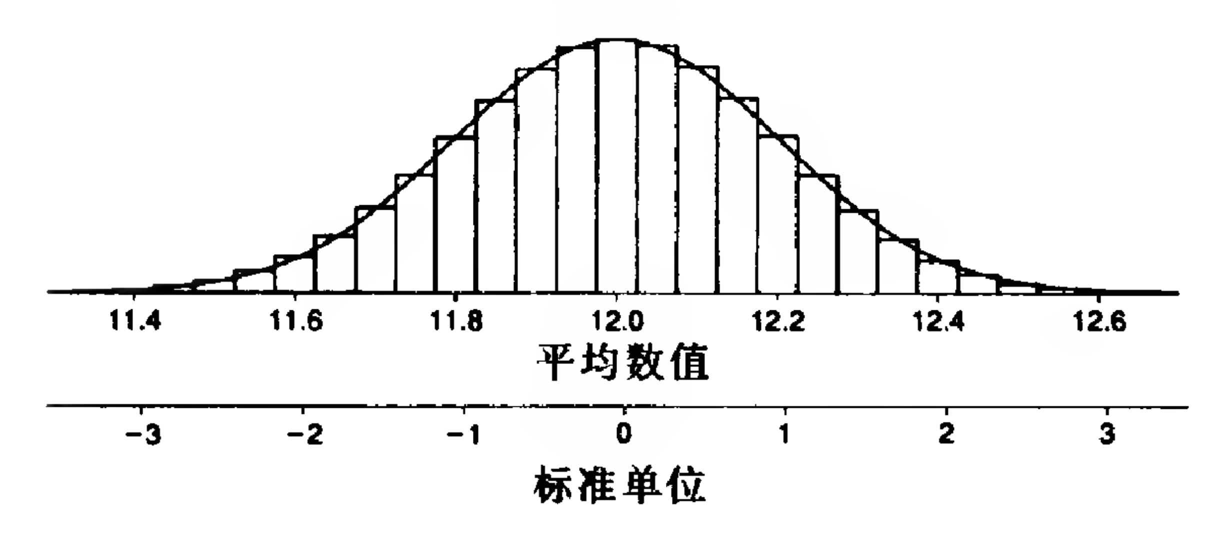
眼下必须从盒中随机不放回地抽取 400 张构成样本。编制程序让计算机来做这件事。第二张长方形图片绘出了 400 个抽得数的直方图。它描述了样本中的 400 人的教育水平的分布。它与第一张图非常相似，尽管受教育 8—9 年的人稍许多了点，那是机会变差。图 2 说明了为什么样本的 SD 是盒子的 SD 的一个好的估计：两张直方图展示了几乎相同的散布量。

至此，我们看到了两张直方图，它们都是关于数据的。现在抽得的平均数的概率直方图登台，这张直方图展示在图 2 的最底部。这第三张直方图不是描述数据的，它描述样本平均数的机会。譬

图2 教育直方图。最上面的一张图展示了全镇25岁和以上的人们教育水平的分布。中间一张图展示了样本中的人们教育水平的分布。这些都是数据的直方图。最下面的一张图展示了从盒子中抽取400个数的平均数的概率直方图;它接近正态曲线。数据直方图的终点约定:以小组区间12—13为例,它包括所有完成12年正规学校教育但没完成第13年的人——没念完大学一年级的中学毕业生。镇的平均数为12.0年,SD为4.0年;样本的相应数值为11.6和4.1







如，取 11.6 和 12.4 年之间概率直方图下的面积为例。这面积表示从盒中抽取 400 个数的平均数将在 11.6 和 12.4 年之间的机会。这面积算得约为 95%。因此，对 95%的抽样组，其样本成员的平均受教育水平将在 11.6 到 12.4 年的范围内。而其余 5%，其样本平均数将在这范围之外。概率直方图下的任何面积可仿此类似解释。

现在你可以看出为什么正态近似是合理的。就象图形所示，即使数据全然不遵循正态曲线，但正态曲线仍是抽得的平均数的概率直方图的很好近似——当样本足够大时。那就是为何正态曲线可用来描述置信水平。（不过，对于小样本，不应该使用正态曲线：第 26 章的第 6 节。）

习题 B

1. 将表 A 中的每一措词与表 B 中的某一个配对。

表 A
总体
总体平均数
样本
样本平均数
样本容量

表 B
抽得数
盒子的平均数
盒子
抽取次数
抽得的平均数

2. 下列每对措词中，有一个有意义，另一个没有。何者为何种？简短地解释之。
   (a)盒子的 SE，盒子的 SD。
   (b)盒子平均数的 SE，抽得的平均数的 SE。
3. 在本节开始的例子中，所有 900 户样本家庭的总收入为 29 163 000 美元。在这个总数中机会误差的可能大小是多少？
4. 在例 3 中，盒子的 SD ________ 4.1 年。用下面选择中的一个填充。

   已知为　　　　由样本估计为

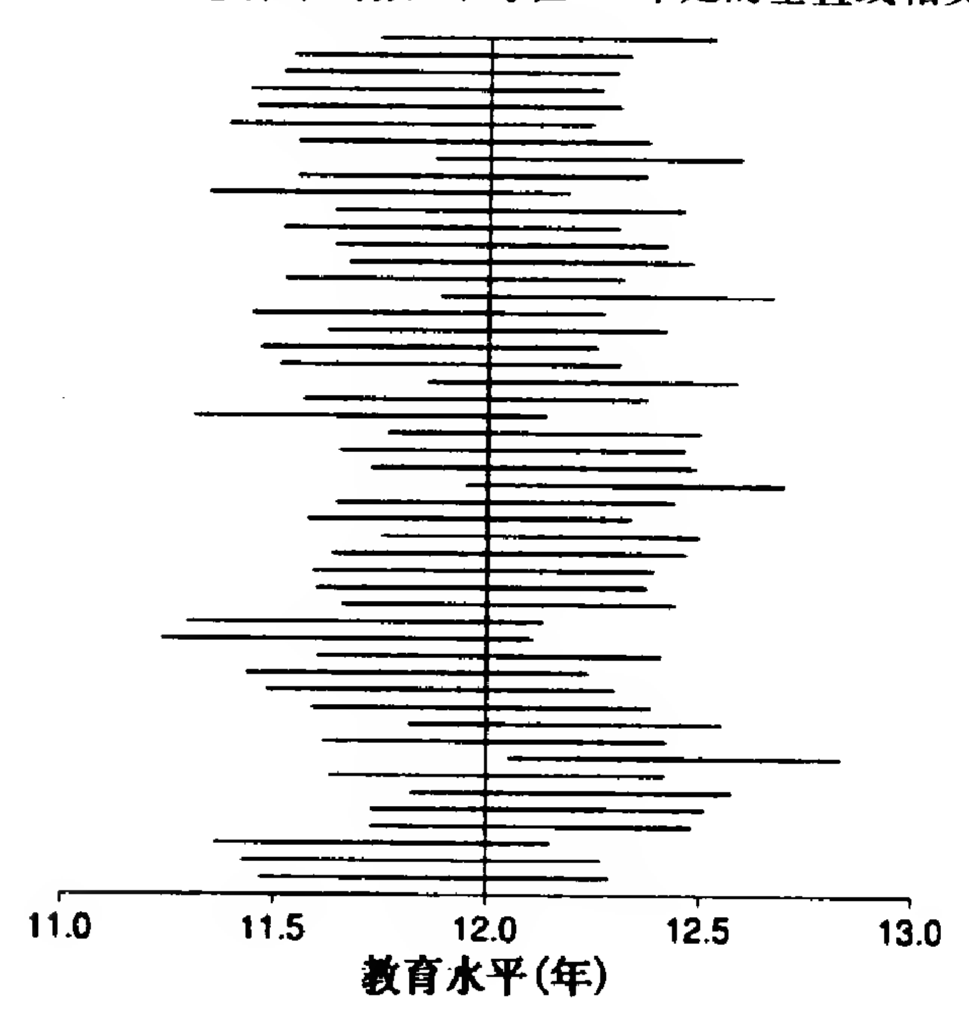
   简短地解释之。
5. 在例 3 中，设 50 个不同的调查机构在镇上各抽取 400 个 25 岁和以上的人为简单随机样本。每个调查机构都得到一个 95%置信区间“样本平均数±2SE”。这些区间中有多少个会覆盖总体平均数？

6. 下图是题 5 中所述试验的计算机模拟。置信区间以不同高度标出以便清楚可见。

(a)为什么这些区间有不同的中心?

(b)为什么它们有不同的长度?

(c)它们中覆盖总体平均数,即与在 12 年处的垂直线相交的,有多少个?



7. 某大学有 30 000 名注册学生。作为调查的一部分,从这些学生中随机抽取 900 人。样本学生的平均年龄算得为 22.3 岁,SD 为 4.5 岁[②]。

(a)全体 30 000 名大学生的平均年龄估计为________,±________左右。

(b)试求全体 30 000 名注册学生平均年龄的近似 95%置信区间。

8. 人口普查局为住房和城市发展部进行年度住房调查[③]。以 1983 年为例,该局估计在美国有 8 千 5 百万住房单元,其中 2 千 9 百万是租赁单元。这些单元的平均租费约为 400 美元,SD 约为 200 美元。

某镇有 10 000 个已被租用的租赁单元。某地方不动产公司对这些单元做了一项调查:随机抽取 400 个单元,并询问了租赁人。在了解的许多问题中,前一个月付的租金是明确的。400 个样本租金的平均数为 368 美元,SD 为 160 美元。对样本租金绘制了直方图,它不遵循正态曲线。

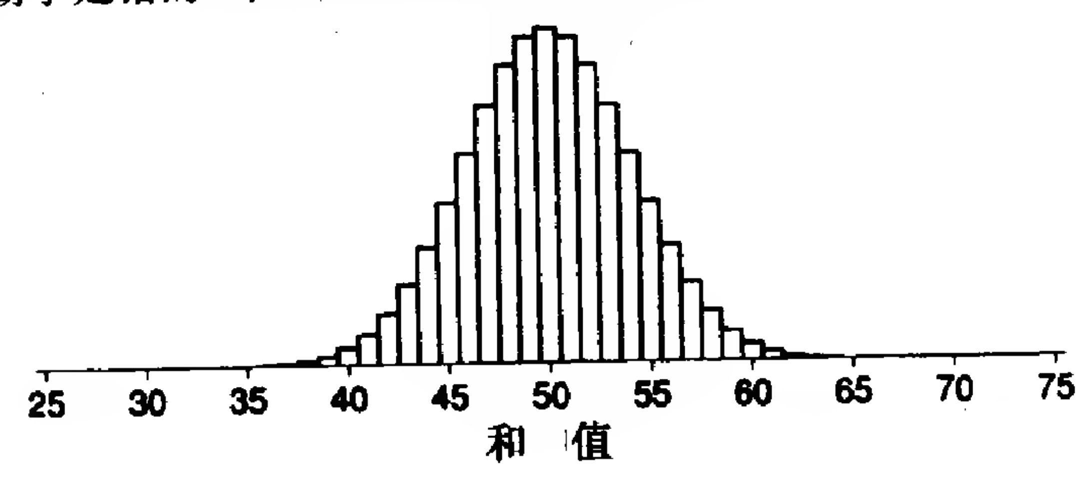
(a)如果可能,试求出这镇上所有 10 000 个已被租用的租赁单元前一个月付的平均租金的近似 68%置信区间。如果这是不可能的,解释为什么

不行。

(b)正确还是错误,并解释之:这镇上所有已被租用的租赁单元中约68%前一个月付的租金在360美元和376美元之间。

9.(续题8.)正确还是错误,并解释之:若随机抽取另外400个已被租用的租赁单元,新样本的平均数约有68%的机会在360美元到376美元范围内。

10.下图是取自盒子 [1] [2] [3] 的25次抽得数之和的概率直方图。但是,某研究人员要在午夜之前用到这些抽得的平均数的概率直方图,他的助手说,“没有什么了不起的,我们必须做的是改变水平轴上的数值。”那样对吗?若对,该助手应将25改成____,将50改成____,以及将55改成____。若该助手是错的,那么需要做什么?解释你的答案。(不需要用到纵向尺度。)



11.(难题)某镇25 000户家庭的人口普查数据是现成的。这25 000户家庭的平均收入为31 700美元,SD为20 000美元。某市场研究公司从这25 000户家庭中抽取了一个400户的简单随机样本。这400户样本家庭的平均收入算得为30 700美元,SD为19 200美元。下图是从这25 000户家庭中随机抽取400户家庭的平均收入的概率直方图。直方图按标准单位绘制。

(a)在这直方图上,标准单位的+1是________。选择:

31 660美元　32 700美元　49 900美元　51 700美元

(b)在标准单位下,30 700美元是________。选择:

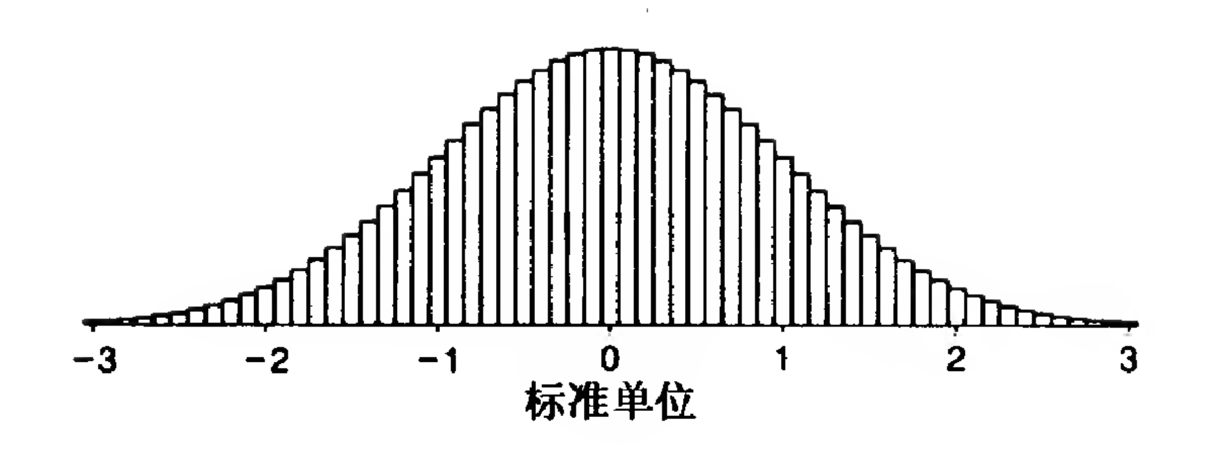
0　　−0.05　　−1　　其它

解释你的答案。

这些习题的答案在第715—717页。

3. 哪一个SE?

SE总是有相同的解释:它是机会误差的可能大小。但是,看来



有许多种SE。何时用何种?最好是写出盒子模型,并决定要对抽得数进行何种算术运算。那将告知你使用哪一个公式,有四种运算供考虑:求抽得数之和,求它们的平均数,分类并计数,或取百分比。相应公式:

$$和的 SE=\sqrt{抽取次数}\times 盒子的 SD$$

$$平均数的 SE=\frac{和的 SE}{抽取次数}$$

$$计数的 SE=来自 0—1 盒子的和的 SE$$

$$百分比的 SE=\frac{计数的 SE}{抽取次数}\times 100\%$$

和的SE 是基础;其它公式全由它导出。这些方式适用于从盒子中随机放回的抽取。

向前或向后推理。当从盒子到抽得数向前推理时,如在第五部分中那样,标准误差可由盒子的成分准确算得。象抽得数的和那样的机会量将在其期望值近旁——但将偏离一个SE 左右。

当从抽得数到盒子向后推理时,你必须依据样本估计盒子的SD。因此SE 本身只是近似的。但是,SE 的解释是类似的。譬如,假设用样本平均数估计盒子平均数,估计将有少许偏离,而SE 指出大约偏离多少。(当样本相当大时,SE 本身中的误差通常会太小而不当一回事。)

SE 表示偏离量的可能大小,它是或正或负的数目。

4. 提 示

本章讲述如何评估根据简单随机样本算得的平均数的精度。由于抽取是随机进行的,因此有可能根据数据中的散布情况和样本的定量来测定精度。这是统计理论的主要成果之一。

对任何数据都可进行算术运算:求出 SD,乘以项数的平方根,然后再除以项数,但是,仅当抽取是随机进行时,上述方法才给出有意义的结果,若数据不是来自正确类型的样本,计算结果可能是无意义的(第 21 章,第 4,5 节,第 22 章,第 5 节)。

提示. 简单随机样本的公式不应该机械地用于其它类型的样本。

习题 C

本习题亦涉及前几章的内容。

1. 某公用事业公司给 50 000 户家庭提供服务。作为顾客观念调查的一部分,他们从这些家庭中抽取了一个 400 户的简单随机样本。样本家庭中拥有电视机的平均台数为 1.86,SD 为 0.90。若可能。试求出全体 50 000 户家庭中电视机的平均台数的 95%置信区间[4]。若这是不可能的,解释为什么不行。
2. 在上题的调查中,400 户样本家庭有 187 户拥有录像机。若可能,试求所有 50 000 户家庭拥有录像机的百分数的近似 99.7%置信区间。若这是不可能的,解释为什么不行。
3. 在前二题的调查中,400 户样本家庭里,399 户至少拥有一台电视机。若可能,试求全体 50 000 户家庭中至少拥有一台电视机的百分数的近似 68%置信区间。若这是不可能的,解释为什么不行。
4. 作为题 1 中所述调查的一部分,访问了 400 户样本家庭中所有 16 岁和以上的人,一共是 900 人。平均来说,调查前的那个星期天样本中人收看电视 5.2 小时,且 SD 为 4.5 小时。若可能,试求该星期天 50 000 户家庭中全体 16 岁和以上的人收看电视平均时数的近似 95%置信区间。若这是不可能的,解释为什么不行。
5. 从一盒中随机放回地抽取 100 次。盒子的平均数为 3.1。

   (a)正确还是错误:抽得的平均数的期望值为 3.1。或这能由给定信息确定

吗？

(b)抽得的平均数的 SE 是多少？或这可由给定信息确定吗？

6. 从一盒中随机不放回地抽取 100 次，抽得数的平均数为 3.1。

(a)正确还是错误：抽得的平均数的期望值为 3.1。或这可由给定信息确定吗？

(b)抽得的平均数的 SE 是多少？或这可由给定信息确定吗？

7. 从盒子中随机放回地抽取 100 次。抽得数之和为 297。你能估计出盒子的平均数吗？你能给你的估计附带标准误差吗？简短地解释之。

8. 盒中装有 250 张票。两个人想估计盒子中数的平均数。他们同意取一个 100 张票的样本，并用样本平均数作为他们的估计。A 想随机不放回地抽取票；B 想取一个简单随机样本。哪一种办法给出更为准确的估计或它有任何差别吗？

这些习题的答案在第 717 页。

5. 复习题

复习题可能包含前面各章的内容。

1. 一盒票有平均数 100，及 SD 20。从盒中随机放回地抽取 400 次。

(a)试估计抽得的平均数落入 80 到 120 范围内的机会。

(b)试估计抽得的平均数落入 99 到 101 范围内的机会。

2. 从一只有 10 000 张票的盒子中随机放回地抽取 400 次。盒子的平均数未知。但是，抽得的平均数为 71.3，且它们的 SD 约为 2.0。正确还是错误，并解释之：

(a)盒中约 68%的票在范围 71.3±0.1 内。

(b)71.3 是盒子平均数的估计，但可能有 0.1 左右的偏离。

(c)盒子平均数的近似 68%置使区间为 71.3±0.1。

3. 某不动产公司想在某有 50 000 户家庭的镇上进行一项调查，以便确定户主上班必须走多远的路程[5]。为此选取了一个 1 000 户家庭的简单随机样本，访问了居住的人，并查明平均来说，样本家庭的户主上班的路程为 8.7 哩，路程的 SD 为 9.0 哩。（所有路程都是指单程；若某户的户主没有工作，上班路程定义为 0。）

(a)该镇全体 50 000 户家庭的户主上班平均路程估计为______
__,这估计值可能偏离________左右。

(b)若可能,试求该镇全体家庭的户主上班路程的平均数的近似 95%置信区间。若这不可能,解释为什么不行。

4.(续题 3.)该不动产公司访问了样本家庭中所有 16 岁和以上的人;有 2 500 名这类人。平均来说,这 2 500 人的上班路程是 7.7 哩,路程的 SD 是 10.2 哩。(同样,若某人没有工作,上班路程定义为 0,又所有路程是单程的。)若可能,试求出该镇所有 16 岁和以上的人上班路程的平均数的近似 95%置信区间。若这不可能,解释为什么不行。

5.(续题 4.)样本家庭中 721 户的户主驾车上班。若可能,试求该镇户主驾车上班的所有家庭的百分数的近似 95%置信区间。若这不可能,解释为什么不行。

6.某年,全美有约 2 700 所高等学校(包括初级大学和社区学院)。作为高等教育持续研究的一部分,Carnegie 委员会从这些高等学校中抽取了一个 400 所的简单随机样本[6]。在这 400 所样本学校中平均注册人数为 3 700。又 SD 为 6 500。委员会估计这 2 700 所高校平均注册学生人数约为 3 700,±325 左右。说出下列每一个陈述是正确的还是错误的,并解释为什么:

(a)估计全美 68%的高等学校注册的学生人数在 3 700－325＝3 375 到 3 700＋325＝4 025 之间。

(b)所有 2 700 所高校的平均注册学生的近似 68%置信区间是从 3 375 到 4 025。

(c)若某人从 2 700 所高校中取一个 400 所的简单随机样本,并从这 400 所样本高校的平均注册学生人数向两侧各移一个 SE,则这个区间将覆盖 2 700 所高校的平均注册学生人数的机会约为 68%。

(d)正态曲线不可能用来描绘置信水平,因为数据不遵循正态曲线。

(e)样本中约 68%的高校的注册学生人数在范围 3 700±6 500 内。

7. (续题 6.)全美高等学校有约 600 000 教师。作为研究的一部分,Carnegie 委员会从这些教师中抽取了一个 2 500 人的简单随机样本[7]。这 2 500 名样本成员在调查前的两年内平均发表了 1.7 篇科研论文,SD 是 2.3 篇。若可能,试求出在调查前的两年内,所有 600 000 名教师发表科研论文的平均篇数的近似 95% 置信区间。若这不可能,解释为什么不行。

8. 某调查机构从有 80 000 户家庭的某城抽取了一个 625 户家庭的简单随机样本。平均来说,每户样本家庭有 2.30 人,SD 为 1.75。试指出下列每一个陈述是正确的还是错误的,并解释之。

(a)样本平均数的 SE 为 0.07。

(b)样本中平均家庭大小的 95%置信区间为从 2.16 到 2.44。

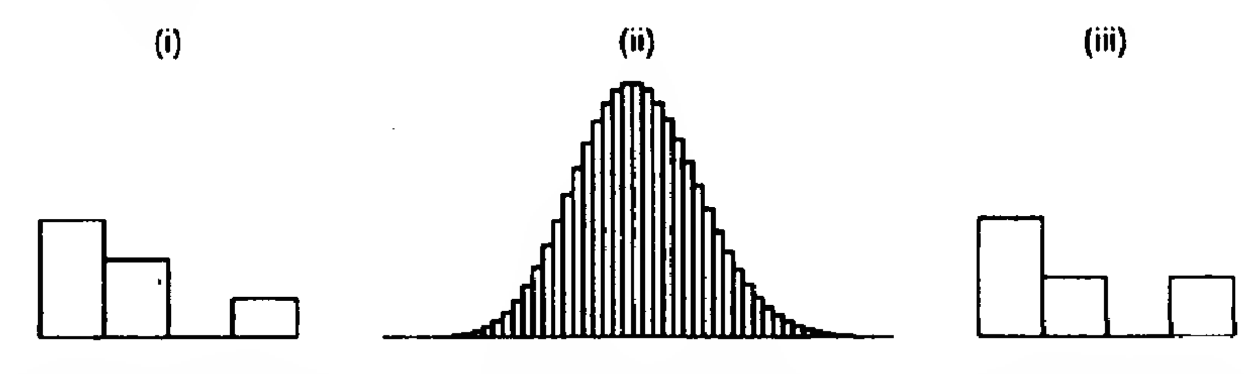
(c)该市平均家庭大小的 95%置信区间为从 2.16 到 2.44。

(d)该市 95%的家庭有 2.16 到 2.44 人。

(e)该市的家庭大小遵循正态曲线。

(f)由于家庭大小遵循正态曲线,故 95%置信水平基本上是正确的。

9. 从盒子 [1] [1] [2] [4] 中随机放回地抽取 25 次。下面直方图中有一张是抽得平均数的概率直方图,另一张是抽得数的直方图,而剩下第三张则是盒子里各数的直方图。哪一幅图配哪一种情况? 解释之。



10. 某学期,加利福尼亚—伯克莱大学有 400 名学生参加统计学 2 的期终考试。他们的成绩平均为百分制的 65.3,SD 为 25。于

是，

$$\sqrt{400}\times 25=500 \qquad 500/400=1.25$$

65.3±2.5 是一个 95%置信区间吗？若是，是关于什么参数的？若不是，为什么不呢？

11. 关于容量为 1 的样本最糟的事是什么？

12. 从盒子中作随机放回抽样。将列在左边的 6 个量与列在右边的它们的值配对。

抽得数之和的期望值	1 161
抽得数之和的 SE	1 200
盒子的平均数	400
抽得的平均数的 SE	40
抽取次数	3
抽得数之和	0.1

并求盒子的 SD。

6. 小 结

1. 当从盒子中随机抽样时，抽得的平均数的期望值等于盒子里的平均数。而抽得的平均数的 SE 等于它们的和的 SE 除以抽取次数。

2. 抽得的平均数可用来估计盒子里的平均数。由于机会误差，估计值将偏离一些，平均数的 SE 告诉你偏离的可能大小。

3. 抽取次数乘以某因子时，抽得的平均数的 SE 除以该因子的平方根。

4. 抽得的平均数的概率直方图将遵循正态曲线，即使盒子里的数并不如此。直方图必须换算成标准单位，而抽取次数必须充分大。

5. 就简单随机样本来说，样本的 SD 用来估计盒子的 SD。盒子平均数的置信区间可通过从抽得的平均数向两侧各移适当多个 SE 求得。置信水平可根据正态曲线读出。这方法仅限于大样本使用。

6. 简单随机样本的公式不应该机械地使用于其它类型的样本。

第七部分　机会模型

24

测量误差模型

根据它所出现的全部，取若干观察值的平均数，将大大地减少所有较小误差的机会，并删除任一大误差的几乎所有可能性：光其结论性思考已足以将这方法的应用不仅推荐给天文学家，而是给涉及做任一种实验的所有其它方面（对它们上述推理是同样适合的）。倘若它们允许在相同情况下重复的话，观察或实验做得越多，结论将越少蒙受误差。

——Thomas Simpson（英国数学家 1710—1761）

1. 估计平均数的精度

在本书的这部分，将使用机会的频率理论来研究测量误差和遗传学。历史上，频率理论被发展用来处理一类非常特殊的问题——在机会游戏中计算可能性。应用这种理论于赌博一类之外的场合是需要作一些努力的。在每一种情况，必须证明被研究的场合与一个象从盒中作抽取的过程相象，对这种情况该理论可以应用。这些盒子模型有时称作机会模型或者随机模型。第一个例子将是测量误差的机会模型；这个主题在第 6 章中介绍过。

简短地回顾一下，任何测量都会有机会误差，如果重复测量，机会误差就会出现少许不同。因此，要得到机会误差的大小，最好是重复测量若干次。这些测量值中的散布程度，由它们的 SD 表示，将给出在单次测量中机会误差可能大小的一个估计。第 6 章就讲到那里。本章将继续讨论，但焦点放在一列测量值的平均数上，不在单次测量上。问题是去估计在这平均数中机会误差的可能大小。如果这些测量就象从盒子中作抽取一样，那末可以使用第五和第六部分的方法。

背景是关于 NB10 的测量。这些测量值以不同的量全都落在不足 10 克之处。第 111 页上的表 1 以单位微克报导了这些不足之量。(一微克是 1 克的百万分之一，约为一粒灰尘的重量。)表中 100 个数的 SD 大约为 6 微克：单次的测量仅仅精确到 6 微克左右。

NB10 重量的最佳估计由所有 100 个测量值的平均数给出，不足 10 克，差 404.6 微克。但是每一个测量值都因误差而产生偏差，所以平均数也不可能完全正确。平均数将比任何单个测量值更精确些，因此它将偏离少于 6 微克。问题在于，在平均数上附加多大的正确允许误差：

$$\text{平均数} \pm \underline{\qquad\qquad} ?$$

答案由平均数的 SE 给出，正如第 23 章中那样，是可以计算的。(计算按盒子模型进行，在下面的第 2 节和第 3 节中讨论。)100 次测量值之和的 SE 可以估计为

$$\sqrt{100} \times 6 \text{ 微克} = 60 \text{ 微克。}$$

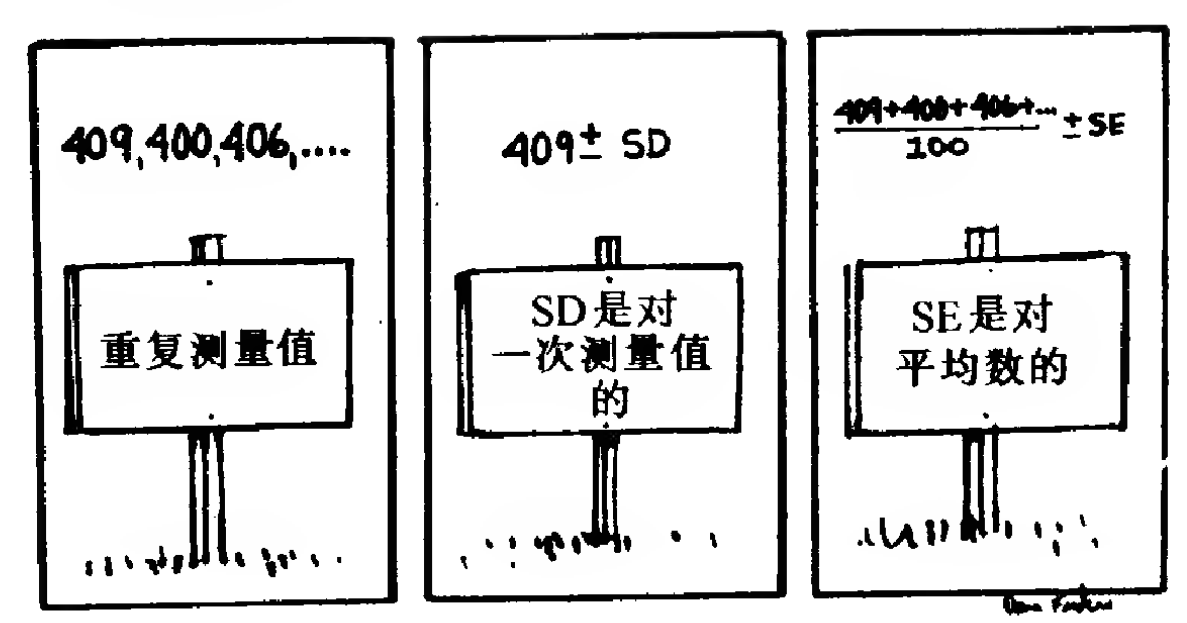
于是 100 次测量值的平均数的 SE 为

$$\frac{60 \text{ 微克}}{100} = 0.6 \text{ 微克。}$$

计算完毕。

在表中所有数的平均数是 404.6 微克。平均数中机会误差的可能大小估计为 0.6 微克。因此 NB10 实际上重 10 克不到 404.6

微克，±0.6 微克左右。



计算中产生两个数：6 微克和 0.6 微克。第 1 个是 100 次测量的 SD，第 2 个是平均数的 SE。两者之间的区别是什么呢?

- SD 说单次测量值精确到 6 微克左右。
- SE 说所有 100 次测量的平均数精确到 0.6 微克左右。

例 1. 对某砝码作 100 次测量。这些测量值的平均数是 1 千克之上 715 微克，SD 是 80 微克。

(a) 单次测量值可能偏离确切重量大约 8 微克，还是 80 微克?

(b) 所有 100 次测量的平均数可能偏离确切重量大约 8 微克，还是 80 微克?

解：单次测量值按一个在大小上与测量值的 SD 类似的机会误差偏离。这是 80 微克。(a) 的答案是 80 微克。对于 (b)，测量值之和的 SE 估计为 $\sqrt{100}\times 80=800$ 微克，因此平均数的 SE 为 800/100=8 微克。那就是 (b) 的答案。

在例 1 中，测量值的平均数的加减数是 8 微克。为了使这更精确些，统计学家使用置信区间，就象在抽样中那样。确切重量的 95% 置信区间可以通过从平均数向两边各移 2 个 SE 而得到。平均数是 1 千克之上 715 微克，2SE 是 2×8=16 微克。因此确切重量以大约 95% 的置信 (confidence) 位于 1 千克之上 699 到 731 微克之间的某个地方。算式为

$$715-16=699, \qquad 715+16=731。$$

再一次，"confidence"(置信)一词在这儿是作为一个提示，提醒我们机会存在于测量过程而不是在被测量的事物中：确切重量不受机会变异的支配。有关抽样情况中类似的讨论，参见第 21 章的第 3 节。

机会存在于测量过程中，而不是在被测量的事物中。

只有当拥有相当多的测量值时，才用正态曲线来获取置信区间。对于少于 25 次的测量值，大多数统计学家将用稍微不同的办法，该方法基于所谓的 t 分布上(第 26 章的第 6 节)。

历史性注。在测量误差理论与霓虹灯广告牌之间存在着这样一个联系。1890 年，人们相信空气是由氮(约 80%)，氧(20%少一点)，二氧化碳，和水蒸气组成的——没有其它东西。化学家能够除去氧，二氧化碳，和水蒸气。剩下的气体应当是纯氮气。Lord Rayleigh 着手比较该剩余气体的重量与等体积的化学纯氮的重量。关于剩余气体的重量的一次测量为 2.31001 克。而纯氮的一次测量值小一点，为 2.29849 克。然而，0.01152 克的差是相当小的，事实上与称重过程所产生的机会误差是可比较的。

这个差可能是由机会误差而引起。若不然，剩余气体里必定包含一些比氮重的东西。Rayleigh 所做的就是重复实验直到他有足够的测量值用以证明来自空气的剩余气体比纯氮重一些。他继续分离出称之为氩的稀少气体，它比纯氮重一些且以很少的量存在于空气之中。后来其他的研究者发现了类似的气体氖，氪，和氙，所有这些都(以微量)自然地存在于空气中。这些气体就是那些使得霓虹灯以不同的颜色闪烁的东西[①]。

习题 A

1. 关于 NB10 的 100 次测量的总数是 40 459 微克。在该总数中机会误差的可能大小是多少？
2. 某些现代化磅秤采用电子荷载元件。重量分布在许多元件上。每个元件将

它所载的重量转换成电流，传播给中心扫描器。这个扫描器将所有电流叠加在一起，并计算出相应的总重量，并将其显示出来。这个过程在一秒钟里重复许多次。因此，一辆满载的棚车(约重 100 000 磅)在穿过一条专门的路径时能被称量，具有仅几百磅大小的机会误差[②]。假定某辆棚车重量的 25 次读数显示出平均数为 82 670 磅和 SD 是 500 磅。该辆棚车的重量估计为________；这个估计可能偏离约________左右。

3. (虚构的。)The British Imperial Yard(大不列颠帝国码)被送往巴黎以 The Meter(米)为参照标准。它的长度被确定了 100 遍。这一列测量值平均为 91.4402 厘米，SD 是 800 微米。(1 微米是 1 米的百万分之一。)
   (a)单次的读数偏离约为 80 微米，还是 800 微米?
   (b)所有 100 次读数的平均数偏离约为 80 微米，还是 800 微米?
   (c)求帝国码确切长度的 95%置信区间。
4. NB10 的确切重量的 95%置信区间是从 10 克之下的 403.4 到 405.8 微克。说出下列每一个陈述是正确还是错误，并解释为什么。
   (a)大约 95%的测量值在这个范围内。
   (b)下一个测量值将在这个范围内约有 95%的机会。
   (c)计量局取 100 次测量值并从其平均数向两边各移 2 个 SE，如是进行约有 95%次成功地覆盖确切重量。
   (d)如果计量局取 NB10 的另外 100 个测量值，大约有 95%的机会使新的平均数将落入 10 克之下 403.4 到 405.8 微克的区间。
5. 取 25 个测量值的平均数时机会误差的可能大小应除以因子 5，10，还是 25?

这些习题的答案在第 718 页上。

2. 机会模型

前节叙述了如何将一个标准误差赋予重复测量的平均数。运算对于任意一列数是容易实施的，但是只有当数据的变异性与从盒中重复抽取的变异性相象时方法才是合理的。

> 如果整个期间数据呈现一定趋势或规律模式，盒子模型不能应用。

理由：从盒子里作抽取不显示整个期间有一定趋势或规律模式。下

面的例子说明如何使用这个想法。

例 2　表 1 给出了从 1790 年到 1980 年美国的人口。这些数看上去象从盒子中作随机抽取的吗?

表 1. 美国的人口,1790—1980 年。

1790	3 929 214
1800	5 308 483
1810	7 239 881
1820	9 638 453
1830	12 866 020
1840	17 069 453
1850	23 191 876
1860	31 443 321
1870	39 818 449
1880	50 155 783
1890	62 947 714
1900	75 994 575
1910	91 972 266
1920	105 710 620
1930	122 775 046
1940	131 669 275
1950	151 325 798
1960	179 323 175
1970	203 302 031
1980	226 545 805

注:常住人口。从 1950 年起,包括 Alaska(阿拉斯加)和 Hawaii(夏威夷)。

来源:Statistical Abstract of the United States. 1988,Tablel .

解:不象。美国的人口数十分稳定地上升。从一个盒子中随机抽取的数并不如此:有时候它们上升而其它时候它们下降。

例 3. 表 1 中的 20 个数平均为 7 千 8 百万,SD 是 6 千 9 百万。一调研人员用下述步骤给平均数赋予一个标准误差:

$$\text{和的 SE} \approx \sqrt{20} \times 69 \text{ 百万} \approx 309 \text{ 百万}$$

$$\text{平均数的 SE} \approx 309/20 \approx 15 \text{ 百万}$$

这合理吗?

解:平均数与 SD 有意义,是描述统计量。它们概括了表 1 中

的部分信息，尽管它们遗漏了相当多：譬如，数值稳定增加这一事实。然而值为1千5百万的SE是无聊的。假如调研人员想知道表中20个数的平均数，它已经被计算出来，没有必要去担心机会误差。

当然，还可能有另外一些东西，譬如表示1790至1980年期间每年的美国人口数的数列的平均数。（关于这列数的每第10个数在表1中给出；在这些数其间的那些数已知精度较差，因为人口普查每十年只进行一次。）于是调研人员作出推断，利用表1的平均数去估计另一个平均数。该估计将偏离一些量。但是平方根法则对于误差的幅度无多大帮助。理由是表1中的数不类同于取自盒子中的抽得数。

平方根法则只适用于取自盒子中的抽得数。

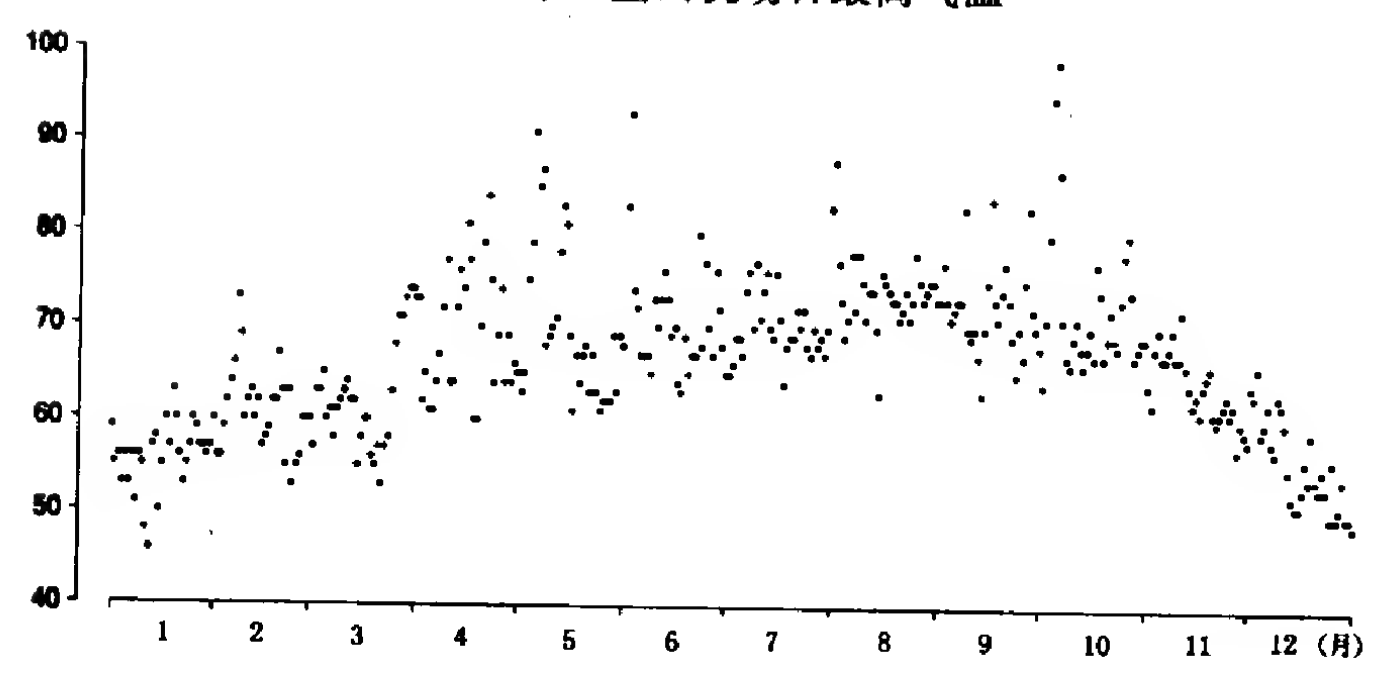
例4. 作一表列，显示San Francisco（旧金山）机场每天的最高温度。这些数据象是从盒子中抽取的吗？

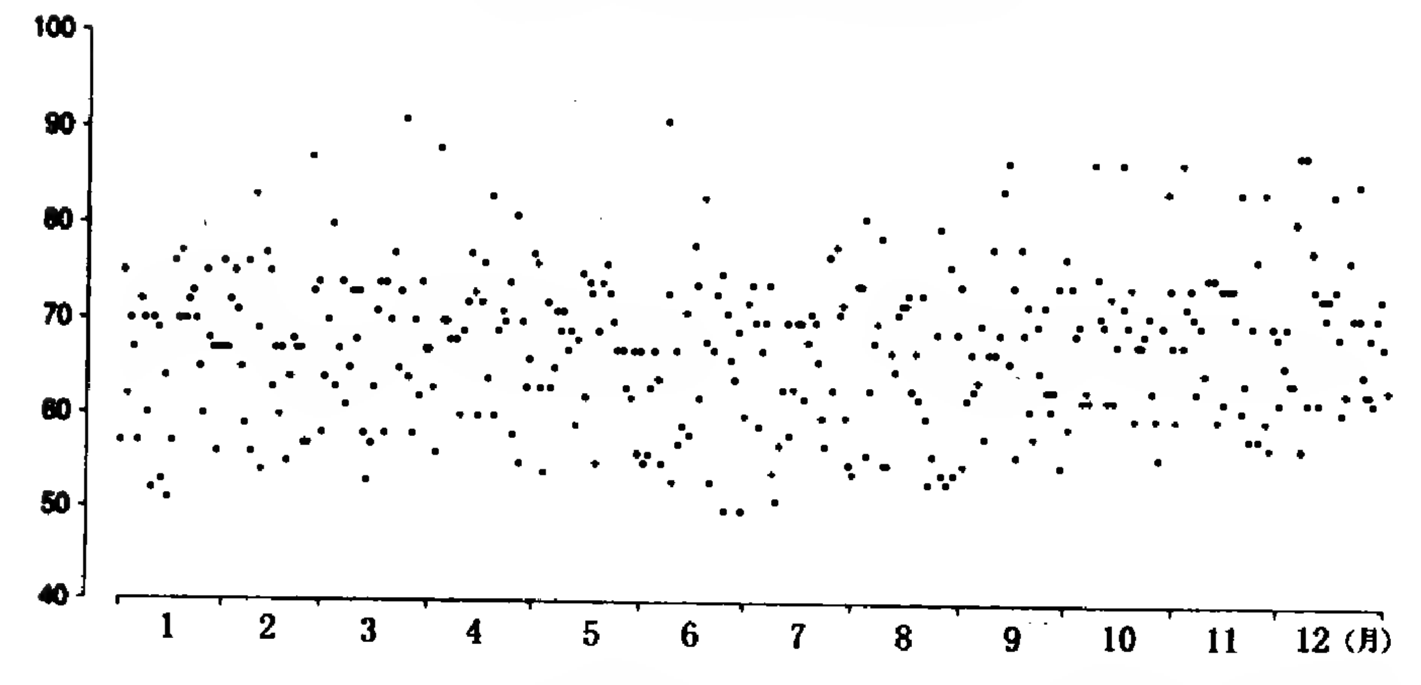
解：不，这些数据有一定的季节性模式：夏天较暖和些，冬天较冷些。数据甚至有其地方色彩，因为在某一天的温度倾向于与以前同一天的温度差不多。

温度数据绘制在图1的上部。对应于1987年中的每一天的上方都有一个点，表示那天的最高温度。季节性模式是明显的：总体来说，夏天的点高于冬天的点。同时，在每一个季节里存在着一个不规则的波状形式。波峰表示一段热的日子——暖和期。寒冷期则在波谷。

作为比较，图1的第二部分是关于一个虚构的机场的气温，在那里平均来说气候象San Francisco（旧金山），但是日最高温度就好象从盒子中抽取的那样。这些数据是随机的：在整个一年内它们没有呈现趋势或模式。对于这类气候，天气预报是无望的。

图 1 温度与盒子模型。第一部分表示 1987 年在 San Francisco (旧金山)机场的日最高温度③。数据有一个季节性模式,夏天比冬天暖和一些。同时,具有地方模式:暖和带与寒冷带。盒子模型不适用。第二部分显示了它们看上去的形式,如果温度是由从盒子中作抽取而产生。



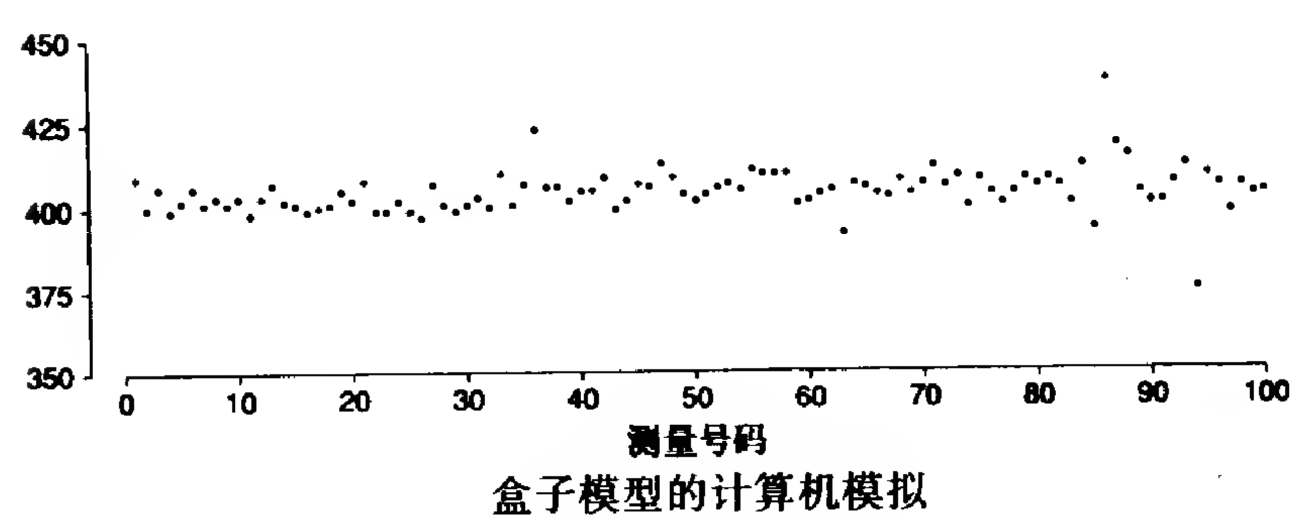


在第 1 节中,我们所做就好象关于 NB10 的测量值是取自盒子的抽得数。这样合理吗? 图 2 中上部是一幅数据图。对每一次测量有一个点与之对应。X－坐标指出它是哪一次测量:第 1 次,或第 2 次,或第 3 次,等等。Y－坐标表示测量值是在 10 克之下多

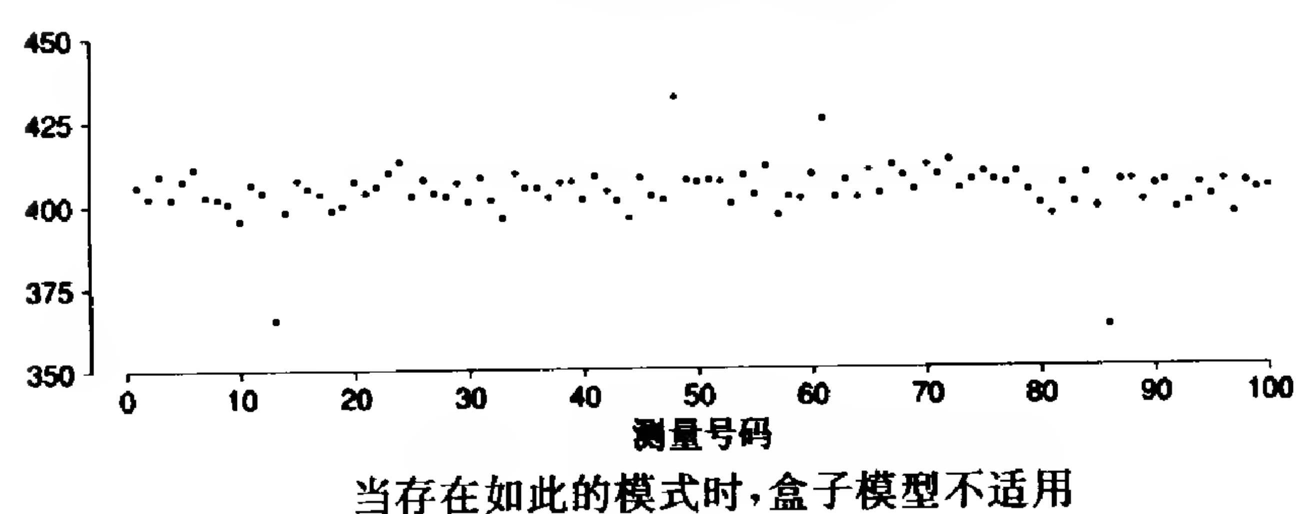
少微克。在整个期间内这些点没有呈现任何趋势或规律模式；它们象取自盒子的抽得数一样随机。事实上，第二部分给出了利用盒子模型在计算机上产生的虚拟数据[④]。如果你不知道哪一幅图是哪一种数据，就很难说出这两个部分之间的不同。作为比较，第三部分（也是由计算机产生的数据）显示出明显的模式，盒子模型将不适用。

图 2 上部绘制了关于 NB10 重复测量值的图象。中间部分对由盒子模型的计算机模拟所产生的虚拟数据绘制图象。这两部分非常类同，显示了盒子模型多么好地描述了实际数据。下部绘制的数据呈现明显的模式：盒子模型将不适用。

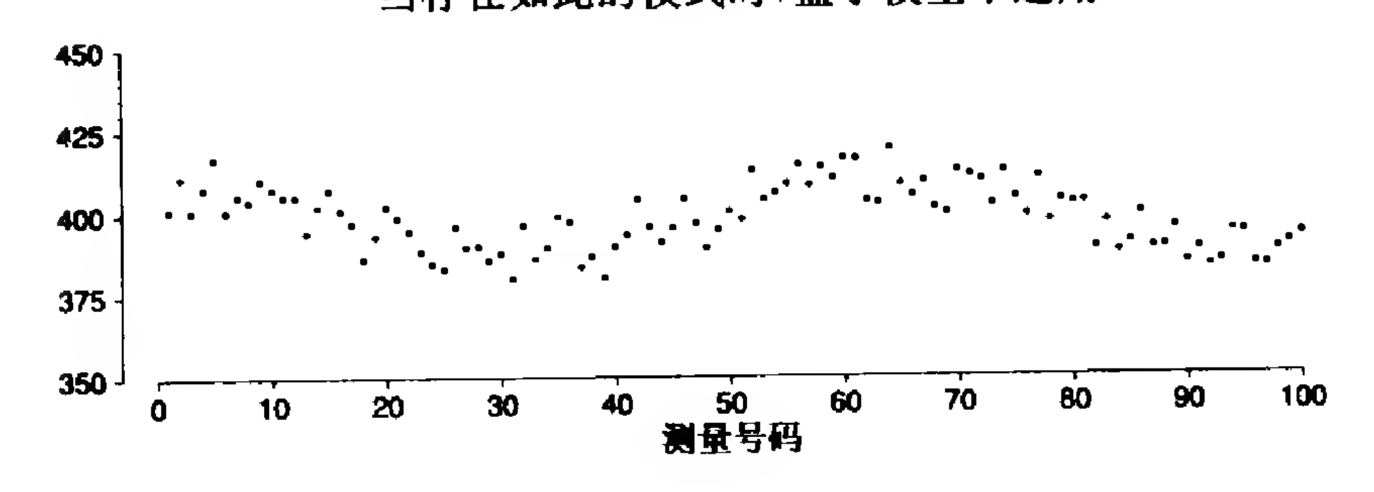
关于 NB10 的测量值



盒子模型的计算机模拟



当存在如此的模式时，盒子模型不适用



关于 NB10 的数据看来象是取自一个盒子中的抽得数并不意外。计量局的研究人员使用如同图 2 上部那样的数据图作为对他们的工作的检验。趋势或规律模式是出了某些差错及需要修整的信号。这个想法对于精密测量工作是基本的——对于批量生产中的质量控制也是基本的，在那儿有缺陷的元件数及时地被标绘出来。

习题 B

1. 一枚图钉抛向空中，着地时要末尖向上要么尖向下。



某人提出如下盒子模型：从盒子 [U] [D] 中作放回抽取，其中 U 表示"尖向上"而 D 表示"尖向下"。另有一人建议采用盒子 [U] [D] [D]。你能怎样确定哪一个盒子较好些？

2. 在 San Francisco（旧金山），平均一年约 17% 的日子下雨。某人对晴雨天序列提出下述机会模型：从一只含有标上"雨"的 1 张卡和标上"晴"的 5 张卡的盒子中作放回抽取。这是一个好模型吗？

3. 某人看完电话薄，并构作了一张显示每个电话号码的最末一位数字的表列。这能模型化为从盒子 [0] [1] [2] [3] [4] [5] [6] [7] [8] [9] 中（放回地）作一系列的抽取吗？对于头一位数字的表列又如何？

4. 某人依次逐个名字地查看完电话薄，并构作了一个表列给出电话薄上每个家庭名字的第 1 个字母。将这个字母序列模型化为从一盒子里作放回抽取有意义吗？（在该盒中有 26 张票，每张票标以字母表的一个字母。）解释之。

5. "精明的职业赌徒，当连续发生 4 次头像时，将赌头像再次发生。已经接连赢 6 次的队将赢 7 次。他相信百分数。业余打赌者将揣测头像不可能再次发生，背面是'应当的。'他将打赌一连串输的队"应当"赢。业余者相信平均数律。"

——Jimmy the Greek, San Francisco Chronicle（旧金山记事）

July2, 1975

(a)由 Jimmy the Greek 所考虑的两种情况：抛一枚硬币和进行竞争性成队运动的比赛之间的主要区别是什么？

(b)抛 Kerrich 的那枚硬币(第 16 章)直到出现接连 4 次落地为头像。这时，假如 Jimmy the Greek 对下一次抛硬币将落地为头像此事开价 5 比 4。(是头像，他付给你 5 美元；是背面，你付给他 4 美元。)你参赌吗？

这些习题的答案在第 718 页上。

3. Gauss(高斯)模型

关于测量误差的盒子模型现在将更详细地描述。基本情况是关于某个量作一系列重复测量。按照模型，每个测量值与确切值相差一个机会误差；这个误差就象从一票盒——误差盒中随机作一次抽取。连串的测量是独立地和在相同条件下进行，因此从误差盒中的抽取是有放回的，为了获取这样一种思想，即机会误差不是系统地取正值或系统地取负值，假设误差盒中的平均数等于 0。这个模型以 Carl Friedrich Gauss(高斯，德国，1777—1855)的名字命名，他做了有关天文学数据中测量误差方面的工作。

在 Gauss 模型中，每测量一次，就从误差盒子里随机放回地抽取一张票。票上的数是机会误差。将它加到确切值上即给出实际测量值。误差盒的平均数等于 0。

在模型中，正是盒子的 SD 给出了机会误差的可能大小。通常，这个 SD 是未知的，必须由数据来估计。例如，取 NB10 的 100 次测量值。根据模型，每个测量值在确切值的附近，偏离为取自误差盒中的一个抽得数：

第 1 次测量值＝确切重量＋取自误差盒的第 1 个抽得数

第 2 次测量值＝确切重量＋取自误差盒的第 2 个抽得数

⋮

第 100 次测量值＝确切重量＋取自误差盒的第 100 个抽得数

对于 NB10 数据，100 个抽得数的 SD 将是误差盒的 SD 的一

个很好的估计[5]。棘手的问题是这些抽得数不能从数据中取得，因为确切重量是未知的。然而，测量值中的变异等于抽得数中的变异，因为确切重量不会随不同次的测量而改变。那就是为什么统计



Carl Friedrich Gauss（高斯，德国，1777—1855）

来自 The Wolff-Leavenworth 收藏，George Arents 研究图书馆，Syracuse 大学。

学家在计算 SE 时采用测量值的 SD 的道理。这就完成了第 1 节计算中的推理[6]。

在应用 Gauss 模型时，可以用一列重复测量值的 SD 来估计误差盒的 SD。当测量值有足够多个时这个估计是好的。

可以有另外的办法得到误差盒的SD。当对于测量过程有大量经验时，较好的是从所有过去的数据去估计SD而不是使用少数几个当前的测量值。理由在于：误差盒属于测量过程，而不属于被测量的东西。

例5.（虚构的。）在对NB10作了数百次测量并求得SD约为6微克之后，计量局的研究人员将这个检验砝码放错了地方。他们出去买了一个新的。用对NB10进行的完全相同的程序在同一磅秤上对它测量。一星期以后，他们积累了25个测量值。它们平均为10克之上605微克，SD是7微克。采取Gauss模型，新砝码是10克之上605微克，加减

6微克　　7微克　　1.2微克　　1.4微克

解：根据模型，每次测量值中的机会误差等同于取自误差盒的一个抽得数。误差盒隶属于磅秤而不是砝码。因此误差盒的SD应该用关于NB10过去大量的数据的SD来估计，而不用关于新砝码的少量当前的数据来估计。因此误差盒的SD估计为6微克。这告诉了在单次测量中机会误差的可能大小。但是25次测量值的平均数中的机会误差大小要小一些。平均数的SE是1.2微克。这就是答案。

最后一点。这里提出的Gauss模型的形式作出了不言而喻的假设：在测量过程中没有偏性。在出现偏性时，每一个测量值是三项之和：

确切值＋偏度＋机会误差

于是平均数的SE不再说测量值的平均数离确切值有多远，而是指出它离

确切值＋偏度

有多远。本章的方法无助于判断偏性。对于NB10的测量，我们不考虑偏性，因为其它的推理方法提出在国家标准计量局的精密称重中的偏性是可忽略不计的。在其它场合，偏性可能比机会误差更严重——而且更难发觉[7]。

习题 C

1. (a)称重一个10克检验砝码。假定Gauss模型没有偏性。如果确切重量是10克之上501微克，且从误差盒中抽取的数是3微克，测量值将为多少？

   (b)如果确切重量为10克之上510微克，从误差盒中抽得的数是−6微克，重复(a)。

2. 关于NB10的第1次测量值是10克之下409微克。假定Gauss模型没有偏性，这个测量值等于

   确切重量＋机会误差。

   你能算出该两项中的每一个数值吗？简短地解释。

3. 在关于NB10的测量值的Gauss模型中，误差盒的SD ________ 6微克。利用下面两组词语中的一个填入空白，并简短地解释之。

   已知为　　　由数据估计为

4. 下图给出了计算机模拟结果：50个假想的研究人员仿效计量局使用的办法去称量NB10。每人取100次测量值并计算平均数，SD，和平均数的SE。图在不同的高度绘制了50个置信区间以使它们能被看清。在模拟中，确切重量取为10克之下405微克。

   (a)为什么这些区间有不同的中心？

   (b)为什么它们有不同的长度？

   (c)应有多少个区间覆盖确切重量？

   (d)实际有多少区间覆盖确切重量？

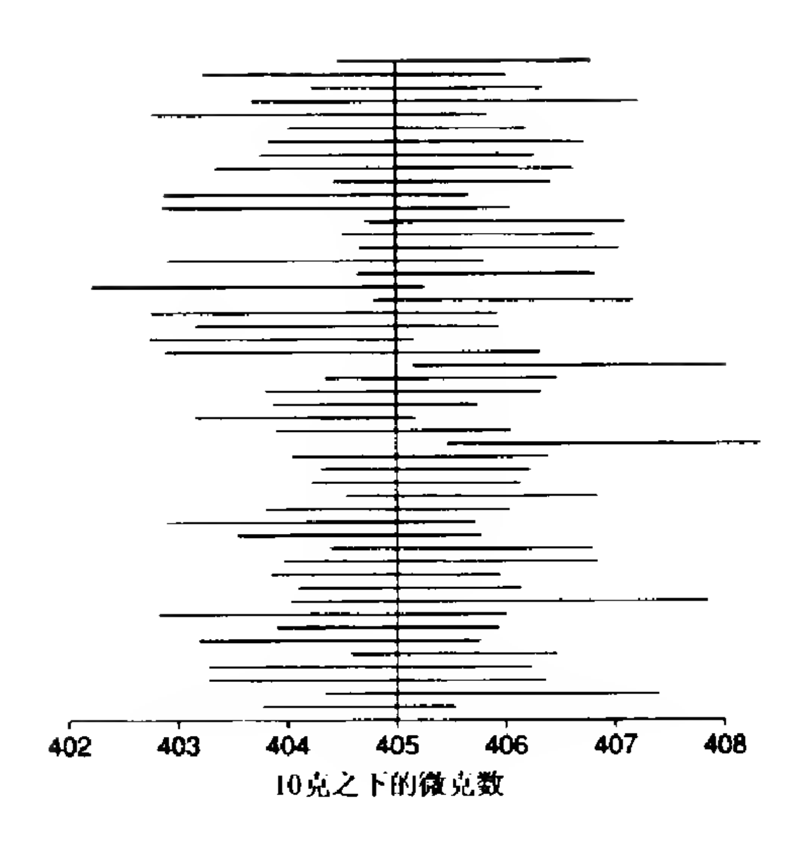
5. 计量局打算称重一个1千克检验砝码100次，并取这些测量值的平均数。他们希望假定为Gauss模型，没有偏性，并且在以往经验的基础上他们估计误差盒的SD为50微克。

   (a)所有100次测量值的平均数可能偏离确切重量约________左右。

   (b)所有100次测量值的SD可能为约________。

   (c)估计所有100次测量值的平均数落在离确切重量左右各10微克之间的范围内的概率。

6. 假如你送一个名义上为10克的砝码到计量局去，要求他们将它称重25次并告诉你其平均数。他们将使用与称重NB10的同样程序，在那里数百次测量的SD大约为6微克。25次测量值的平均为10克之上307微克，SD



大约为 5 微克。你的砝码是 10 克之上 307 微克，加减约：

5 微克　　6 微克　　1 微克　　1.2 微克

（你可以假定为 Gauss 模型，没有偏性。）

7. 关于光速作了 25 次测量。它们平均为 300 007 和 SD 为 10，单位是每秒千米。在(a)中填写空白，然后说出(b—f)中的每一个是正确还是错误；简短地解释你的答案。

（你可以假定为 Gauss 模型，不具有偏性。）

(a)光速估计为________；这个估计可能会偏离________左右。

(b)每次测量值偏离 300 007 约 10 左右。

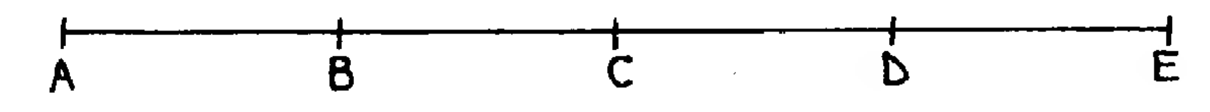
(c)全部 25 次测量值平均数偏离 300 007 约 2 左右。

(d)如果做第 26 次测量，它将偏离光速的确切值约 2 左右。

(e)光速的 95％置信区间为 300 007±4。

(f)25 次测量值的平均数的 95％置信区间是 300 007±4。

8. 一测量员正在测量 5 个点 A，B，C，D，E 之间的距离。它们全都在一条直线上。她发现 4 个距离 AB，BC，CD，和 DE 中的每一个测量为 1 哩±1 吋左右。这 4 次测量是用相同的方法独立地完成的。



从 A 到 E 的距离大约为 4 哩，加减约

4 吋　　2 吋　　1 吋　　1/2 吋　　1/4 吋

简短地解释之。(你可以假定为 Gauss 模型,不具有偏性。)

9. 测量误差的概念常常被用到心理测试的结果上。公式为

实际测试分=真正测试分+机会误差。

机会误差这一项反映了偶然因素,象被测对象的情绪,或者运气。你认为 Gauss 模型适用吗?

这些习题的答案在第 718—719 页上

4. 结 论

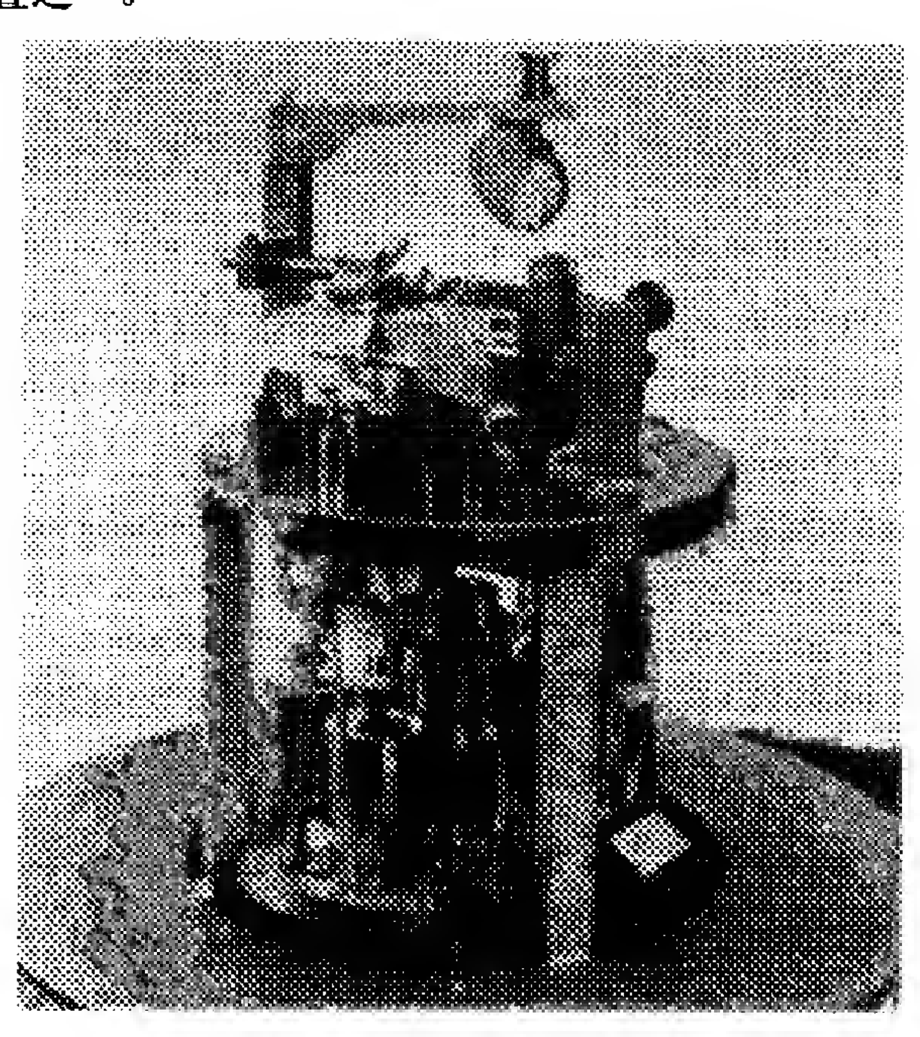
NB10 只是一块金属。在平台,齿轮,和杠杆组成的奇妙装置上称重。这些称重的结果经受了包括标准误差,正态曲线和置信区间在内的统计分析。正是 Gauss 模型将数学与 NB10 联系起来。机会误差等同于取自盒中的抽得数,因此它们的平均数等同于这些抽得数的平均数。抽取次数是如此地大以使得平均数的概率直方图将非常接近地拟合正态曲线。没有模型就没有盒子,没有标准误差,以及没有置信水平。

统计推断是指利用机会方法从数据中得出结论。对平均数赋予一个标准误差就是一个例子。现今总可以用机器操作做完 SE 程序。许多计算机程序将为你做这件事。甚至可以将输出结果标明为"标准误差"。不用着迷于算术或者专门术语。方法仅平方根法则下有意义。含蓄的假设称数据等同于从盒子中抽取的结果;这是个老观点,但值得重提。许多研究人员对假设不给予注意:其结果是,他们报告的"标准误差"经常是无意义的。

统计推断能够通过对数据提出一个明确的机会模型而证明其合理性。没有盒子,就没有推断。

第二部分和第三部分聚焦于描述性统计:绘图或计算以概括数据和明白地表示出显著的特征。这样的技巧可以相当普遍地使用,因为它们不包含任何有关数据从何处来的隐蔽的假设。然而,对于统计推断来说,模型是基本的。

图 3 建在国家标准局的精密称重装置。这个装置是供称重 1 千克的，且送往在巴黎的国际度量衡局。这是迄今建造的最精密的称重装置之一。



5. 复习题

1. 新的激光测高计能测量高度在数吋之内，而无偏性，且测量值没有趋势或规律模式。作为实验的一部分，对某一山峰的高度取得 25 个读数。它们平均为 81 411 吋，它们的 SD 是 30 吋。填(a)中的空格，然后指出(b－f)中每一个是正确还是错误；简短地解释你的答案。(你可以假定为 Gauss 模型，不具有偏性。)

   (a)山峰的高度估计为________；这个估计可能偏离________左右。

   (b)81 411±12 吋是 25 个读数的平均数的 95％置信区间。

   (c)81 411±12 吋是山峰高度的 95％置信区间。

   (d)下一次读数大约有 95％的机会将落在 81 411±12 吋的范围内。

(e)大约95%的读数落在81 411±12 吋这个范围内。

(f)如果测得另外25个读数，大约有95%的机会它们的平均数将落在81 411±12 吋的范围内。

2. 测量光速25次。平均数为每秒299 774千米，SD是每秒14千米[8]。假定为Gauss模型，不具有偏性。求光速的近似95%置信区间。

3. 在习题2中，测定光通过一个确定距离时所需时间。距离被测量57次，这些测量值的平均数是1 594 265毫米。你还需要知道什么以确定这个值有多精确?

4. 习题3表明了在习题2中描述的测量中一个可能的偏性来源。它是什么?

5. 1987年，San Francisco (旧金山)机场的日最高温度的平均数是66.6度，SD是8.4度(图1)。于是

$$\sqrt{365}\times 8.4\approx 160 \text{ 度}, 160/365\approx 0.4 \text{ 度}。$$

正确还是错误:San Francisco (旧金山)机场平均日最高温度的一个近似95%置信区间是66.6±0.8度。简短地解释之。

6. 校准实验室用相同的程序测量1千克检验砝码数年之久。他们积累了数百个测量值。这些测量值的SD是18微克。现在某人送来一个1千克砝码要求用相同的程序校准。实验室对新的砝码做了50次测量，其平均数在1千克之上78.1微克，它们的SD是20微克。若可能的话，求这个新砝码的值的一个近似95%置信区间。(你可以假定为Gauss模型，不具有偏性。)

7. 在一长系列试验中，发现平均要花CPU时间58秒以实施一计算机程序，其SD是2秒。在数据中不存在趋势或规律模式。将花费CPU时间约____秒±____秒左右，以实施程序100遍。(CPU是"中心处理装置"，那里的机器执行逻辑和算术。)

8. 一架机器制作平均重量为4.0盎司的黄油棒;重量的SD是0.05盎司。数据不存在趋势或规律模式。在一包中有4根黄油棒。

(a)一包重________±________左右。

(b)一商店买 100 包。估计他们得到 100 磅黄油，误差在 2 盎司之内的机会。

（译者注：1 磅＝16 盎司。）

9. 对这个问题的两个部分你可以假设为 Gauss 模型。指出每一个断言是正确的还是错误的，并给出理由。

(a)如果你只有一个测量值，你不能够估计在它之中的机会误差的可能大小——你必须取另一个测量值；并且看看它变化了多少。

(b)如果你所有的是 100 个测量值，你不能够估计它们的平均数之中的机会误差的可能大小——你必须取另 100 个测量值，并且看看平均数变化了多少。

10. 正确还是错误并解释："如果数据不遵循正态曲线，你不能利用正态曲线去得到置信水平。"

11. "所有的测量进行两次。假如两个职员参加，由不同的人进行成对测量。为了使总误差达到最小，差异大于某确定的范围就进行第 3 次测量，并且倘若需要的话就做第 4 次，直到所得的 2 个测量值在允许的设置范围内。在不一致场合，由测量员决定这 3 或 4 个结果中哪一个是最有代表性的并指定它为统计记录的内容。在令人满意的一致场合，统计记录按惯例以第 1 次测量记录为基础。"简短地评论之[9]。

6. 小 结

1. 根据关于测量误差的Gauss（高斯）模型，每作一次测量，就从误差盒中随机放回地抽取一张票。票上的数是机会误差。它加在被测物的确切值上，给出了实际测量值。误差盒的平均数等于0。这儿，假定偏差是可忽略不计的。

2. 当应用 Gauss 模型时，许多重复测量值的 SD 是误差盒 SD 的一个估计。它告诉了在一次测量中机会误差的可能大小。

3. 系列的平均数比起任意一次测量更精确些，相差一个测量次数的平方根这个因子。这假定了数据遵循 Gauss 模型。

4. 关于被测物的确切值的近似置信区间可以通过从测量值的平均数向两边各移适当个数的 SE 而求得；置信水平是从正态曲线获得的。只要盒子模型适用且有足够多的测量值，该近似是好的。

5. 对于 Gauss 模型，机会变异存在于测量过程而不是在被测物本身。"confidence"（置信）一词提醒你这一点。

6. 如果模型不适用，那么获得置信区间的办法也不能使用。特别地，如果数据中存在任何趋势或规律模式，公式可能给出愚蠢无聊的答案。

7. 统计推断利用有关数据的明确的机会模型而证明其合理性。

25

遗传学中的机会模型

我决不相信上帝与世界万物玩掷骰子游戏。

——Albert Einstein (1879—1955)

1. Mendel(孟德尔)如何发现基因

本章很难,可以跳过不读而不会丧失本书中主题的头绪。将它收入是出于两个理由:

- Mendel(孟德尔)的遗传学理论是伟大的科学;
- 该理论揭示了在人类活动中简单机会模型的力量。

1865 年,Gregor Mendel 发表一篇文章对遗传提供了科学的解释,并且最终在人类对生命的认识上引发了一场革命[①]。由于命运的荒谬且意想不到的转折,这篇文章被忽视了大约 30 年,直到该理论被德国的 Correns,荷兰的 de Vries,和奥地利的 Tschermak 等三人同时重新发现。现在 De Vries 和 Tschermak 被认为在他们发表文章之前已经看过 Mendel 的论文,但是 Correns 明显地是自己发现这种思想的。

Mendel 的实验全都是用普通的豌豆实施;这儿给出这些实验之一的简短报道。豌豆种子要末是黄色的要末是绿色的[②],且一株作物能够结出两种颜色的种子。Mendel 培植纯粹黄色的品系,就

是说，这个品系中每一代的每一个植物只有黄色的种子；独立地他又培植纯粹绿色的品系。然后他使纯黄色品系植物与纯绿色品系植物杂交。譬如，他用来自黄色品系的花粉使绿色品系植物的胚珠受粉。（另一种方法，用绿色品系的花粉使黄色品系植物受粉，完全得出相同结果。）黄-绿杂交产生的种子，以及它们长出的植物，称为第一代杂交。第一代杂交种子全是黄色的，与纯黄色品系的种子不可区别。这样绿色似乎完全消失了。



Gregor Mendel（奥地利，1822—1884）
选自布尔诺，摩拉维亚博物馆收藏品。

这些第一代杂交种子生长成为第一代杂交作物，Mendel 使它们自身杂交，产生了第二代杂交种子。这些第二代杂交种子的某一些是黄色的，而另一些是绿色的。这样对于第一代绿色消失了，而在第二代中又重新出现。甚至更令人惊异的，绿色以确定的、简单的比例重新出现：在第二代杂交种子中，大约 75％为黄色和 25％

为绿色。

在这个规律性后面是什么呢？为了解释它，Mendel 假设了实体，现今称之为基因[3]。按照 Mendel 的理论，存在两种不同的基因变异体，它们成对地发生以控制种子颜色。在这里它们被表示为 y（代表黄色）和 g（代表绿色）。因此，存在四种不同的基因对：y/y，y/g，g/y，和 g/g。基因对按下述规则控制种子颜色：

- y/y，y/g，和 g/y 产生黄色
- g/g 产生绿色

正如遗传学家所说，y 是显性性状，而 g 是隐性性状。正是种子中的基因对——不是母本——确定了种子将是什么颜色，组成种子的所有细胞含有相同的颜色基因对。这样完成了模型的第一部分。

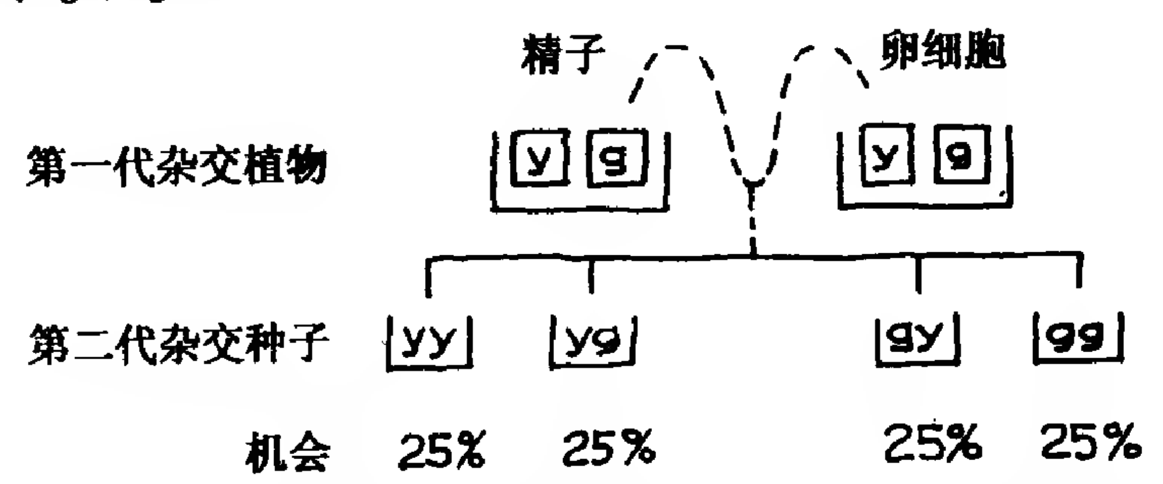
现在种子生长且成为植物；该植物中的所有细胞也带着种子的颜色基因对——有一个例外。性细胞，不论是精子还是卵细胞，只包含基因对中的一个基因[4]。例如，原来的细胞包含基因对 y/y 的植物将产生精子细胞每一个含有基因 y；同样地，它将产生卵细胞每一个含有基因 y。另一方面，原来细胞含有基因对 y/g 的植物将产生一些包含基因 y 的精子细胞，和一些包含基因 g 的精子细胞。事实上，它的一半精子细胞包含 y，而另一半包含 g，同样地，它的一半卵细胞包含 y，而另一半卵细胞包含 g。

这个模型说明了实验结果的原因。纯黄色品系的植物有颜色基因对 y/y，因此精子和卵细胞全都只含基因 y；同样地，纯绿品系的植物有基因对 g/g，因此它们的花粉和胚珠只含有基因 g。使一个纯黄色品系植物与一个纯绿色品系植物杂交，譬如，相当于用一个 y—精子使 g—卵细胞受精，产生了一个带有基因对 y/g 的受精细胞。这个细胞繁殖本身，最终成为一个种子，在那里所有的细胞带有基因对 y/g 且颜色是黄色的。该模型解释了为什么所有第一代杂交种子都是黄色的，没有一个是绿色的。

第二代怎么样？这个嘛，第一代杂交种子成长变为带有基因对 y/g 的第一代杂交植物。在适当的时候，这个植物产生了精子细

胞，其中一半将含有基因 y 而另一半含有基因 g；它也产生了卵细胞，其中一半含有 y 而另一半含有 g。当两个第一代杂交植物交配时，每一个产生第二代杂交种子的受精细胞从每一个亲本中随机地得到一个基因——因为它是由一个精子细胞和一个卵细胞的随机组合而形成。根据种子的观点，就好象从两个盒子（或亲本）中的每一个随机地选取一张票（或基因）：在每一个盒子中，一半票标记 y，另一半票标记 g。见图 1。

图 1 关于种子颜色的遗传确定的 Mendel 机会模型：从每一个亲本中随机地选取一个基因。给出了每一种组合的机会。首先列举了精子基因，但是在受精之后，根据种子的颜色是不可能区分组合 y/g 和 g/y 的。



结果是，种子具有 25%的机会得到含两个 g 的基因对，从而是绿色的；有 75%的机会得到含有一个或二个 y 的基因对，从而是黄色的。种子数与花粉数相比很小，因此对各种各样的种子的挑选基本上是独立的。结论：第二代杂交种子的颜色将如此来确定，就好象从盒子 | 黄色 黄色 黄色 绿色 | 中进行一系列有放回的抽取。那就是模型如何去说明在第二代中大约有 25%的种子再出现绿色的原因。

Mendel 从他的实验现象到他的理论结果真正地作出了一个非常大胆的跳跃。他那关于遗传链的重建完全基于这里所讨论的一类统计依据。可以说他是正确的。他预料存在的过程的大多数现今可以照相。遗传学的现代研究和分子生物学正在揭示遗传的化学基础，并且提供了关于 Mendel 假设实体的存在性的充分直

接的证明。正如我们今天所知道的，基因是染色体上DNA(脱氧核糖核酸)的切片——下一页上图2中的黑斑。

基本上遗传的同样原理在生命的所有形式中起作用，从海豚到果蝇。因此由Mendel提出的遗传模型揭示了生命的重大秘密之一。一粒豌豆种子总是长出豌豆，决不会是西红柿，或者是鲸鱼，这又怎样呢？尽管有本章开头Einstein(爱因斯坦)的引语。这个答案证实了决定性状况中包含有机会。

习题A

1. 在一些实验中，第一代杂交豌豆回头与一亲本交配。如果y/g作物与g/g作物交配的话，大约有多少百分数的种子是黄色的？在1 600粒这样的种子中，超过850粒是黄色的机会是多少？
2. 金鱼草的花色由一基因对所控制。有两个基因变体，r(表示红色)和w(表示白色)。规则是：

   r/r 造成红花

   r/w 和 w/r 造成粉红色花

   w/w 造成白花

   这样r和w都不是显性状态的。它们的效应是加性的，就象混合的红颜料与白颜料。

   (a)算出由下述杂交：白×红，白×粉红，粉红×粉红所产生的红，粉红，和白花作物的期望百分数。

   (b)在来自于粉红×粉红杂交的400株植物中，有190和210株之间具粉红色花的机会是多少
3. 金鱼草叶具三种宽度：宽，中等，和窄。在培育试验中，得到下述结果：

   宽×宽→100%宽

   宽×中→50%宽，50%中

   宽×窄→100%中

   中×中→25%窄，50%中，25%宽

   (a)你能构作一个遗传模型来解释这些结果吗？

   (b)从下述杂交中的每一种，你能指望何种结果？

   窄×窄，窄×中。

图 2 显微照相。这些细胞来自于豌豆作物的根稍，被放大了约 2000 倍。在中心处所示的细胞即将分离。在这个阶段，每个单个染色体由两个紧挨在一起的完全相同的片组成。有 14 个染色体排列成 7 个同质体的对，用罗马数码Ⅰ至Ⅶ标志。控制种子颜色的基因对位于染体对Ⅰ，每一个染色体上的基因之一[5]



来源：纽约州农业实验站，Geneva，N. Y。

4. 人类的眼睛颜色是由一个基因对确定，它具有棕色的显性状态和蓝色的隐性状态。在某家庭中，丈夫有一个蓝眼睛的父亲；他本人有棕色的眼睛。妻子有蓝眼睛。他们计划要有 3 个孩子。所有 3 个孩子将有棕色眼睛的机会是多少？（最好是精确地算出它而不是应用正态近似。）

这些习题的答案在第 719—720 页上。

2. Mendel（孟德尔）的论据符合他的模型吗？

Mendel 的发现列为最伟大的科学之一。如今，他的理论不仅得到充分地证明而且是极其有权威的。可是他本人的实验论证有多好呢？Mendel 的数据资料证实了他的理论吗？R. A. Fisher 回答说，非常好：

> …在 Mendel 的预期和他的报告结果之间相一致的总水平证实了这比起在最好的数千次重复中所能指望的还要接近一些。数据已经明显地被故意篡改，在检查了各种各样的可能性之后，我毫不怀疑 Mendel 被一个园艺助手所欺骗，这助手非常了解对所做的每个试验他的委托人指望的是什么[6]。

换句话说，撇开园丁的话，Fisher 是说 Mendel 的数据是捏造的。这个断定的基础是 Mendel 观察到的频率很不自在地接近于他的期望频率，比正常的机会变异所允许的接近得多。譬如，在一项实验中，Mendel 得到 8 023 颗第二代杂交种子。他预料它们有 $\frac{1}{4}\times 8\ 023\approx 2\ 006$ 颗是绿色的，而观察到 2 001 颗是绿色的，差异是 5。根据他自己的机会模型，那就象从匣子

黄色	黄色	黄色	绿色

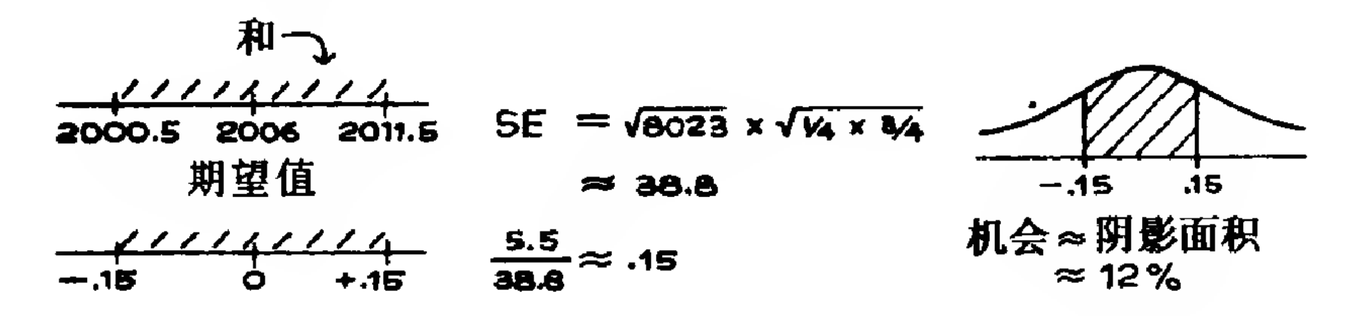
中有放回地抽取 8 023 次。在这个模型中，在绿色的数目与所预料的数目之间观察差异在 5 或 5 以下的机会是多大呢？换句话说，绿色的数目是在

$$\tfrac{1}{4}\times 8\ 023-5\approx 2\ 001 \text{ 和 } \tfrac{1}{4}\times 8\ 023+5\approx 2\ 011$$

之间的概率是多大？这就象从匣子

0	0	0	1

中有放回地抽 8 023 次并询问其和在 2 001 与 2 011 之间，包括这两个数在内的机会。这个机会可以使用正态近似来估计，但保持矩形的边界，如第 347—350 页那样。



就是说，大约 88%的时机，机会变异将使 Mendel 的预料值与他的观察值之间的差异大于他所报告的。

单是这个论据，本身并不令人困惑。麻烦在于，Mendel 实验的每一个（有一个例外将在下面讨论）都显示了期望与观察之间的这种不寻常的接近一致。用 χ^2—检验处理这些结果（第 28 章），Fisher 证明了接近于 Mendel 所报告的那种一致的机会大约是 100 000 分之 4。换个方式，假设有成百万科学家正忙于重复 Mendel 的实验。设想对每位科学家，都计量他所观察到的频率与按 χ^2 统计量所期望的频率之间的差异。又假设 Mendel 的机会模型是绝对正确的。那末根据机会法则，那些假想的科学家每 100 000 人中大约有 99 996 人报告的观察值与期望值之间的差异将大于 Mendel 所报告的。因此存在两种可能性：

· 要末 Mendel 的数据是经过加工的

· 要末他相当走运。

第一个可能性较易于接受。

Fisher 论据有一个方面值得较密切的注意：讨论是技术性的，读者可以跳过而到下一节的开头。Mendel 是以六个不同于种子颜色的特征着手的；譬如，它们中之一是豆荚的形状，或者胀大了（显性状态形式），或者缩小了（隐性状态形式）。遗传机制十分类同

于种子颜色：豆荚形状是由一个基因对所控制。形状基因有两个变量，表示为 i(使胀大的)和 c(使缩小的)。基因 i 是显性状态，因此 i/i 或者 i/c 或者 c/i 使豆荚胀大，而 c/c 使豆荚缩小。控制种子颜色的基因对独立于控制豆荚形状的基因对起作用。

在种子颜色和豆荚形状之间有一个差别。豆荚形状是亲本植物的一个特征，它完全不受受精花粉的影响。于是，如果显示种子豆荚的隐性状态缩小形式的纯种中的一株植物使用来自显示显性状态胀大形式的纯种植物的花粉受精的话，所有结出的种子豆荚将具有隐性状态缩小形式。但是当这样杂交的种子生长为成熟的第一代杂交植物，并结出它们自己的种子豆荚，它们都将呈现显性状态胀大形式。

如果第一代杂交互相交配，第二代杂交植物中大约 3/4 将呈现显性状态形式而¼为隐性状态形式。如图 1 所示，带有显性状态胀大形式的第二代杂交植物中，大约

$$\frac{25\%}{25\%+25\%+25\%}=\frac{1}{3}$$

应该是 i/i 而另外⅔应该是 i/c 或 c/i。Mendel 在 600 株植物中检查出这一点，发现了 201 株 i/i，一个聊以自慰的过分接近于期望值 200 的结果[7](如此接近的一致，其机会仅约为 10%)。

可是更坏的还在后头。你不可能仅仅通过观察将 i/i 同其他的分辨开来；外貌是一样的。因而 Mendel 怎样将它们分类呢？好，如果博物学家不干扰的话，豌豆将自己传花粉。因此 Mendel 取他的显示显性状态胀大形式的第二代杂交植物，并且随机地选择了 600 株。然后从他所选出的植物中的每一株培育了 10 颗后代。如果植物纯种培育而且所有 10 个后代均表现出显性状态胀大形式，他就将它归为 i/i 类；而如果植物产生表现隐性状态缩小形式的任何后代，他就将它归为 i/c 或 c/i 类。

这个方案存在一个困难，Mendel 似乎忽略了。如图 1 所示，一个自身受精的 i/c 的后代将至少包含一个显性状态基因 i，从而

表现出显性胀大形式，这样的机会是$3/4$。因此 i/c 的 10 个后代自身杂交将都显示显性状态形式的机会是$(3/4)^{10}$。那是一个相当大的机会，大约为 6%。对于 c/i 同样如此。因此植物归类为 i/i 的期望频数比 200 大一点，因为 400 个 i/c 和 c/i 的约 6%将不正确地归类为 i/i。将这个归错类的机会一起算入，植物归类为 i/i，不管是正确地还是不正确地，其期望频数是

$$200+0.06\times400=224。$$

Mendel 观察的频数(201 株归类为 i/i)离这个修正的期望值相当地远。如此大的差异的机会大约是 5%。如 Fisher 所断定的："不存在摆脱困境的轻松方法。"

3. 回归律

第三部分讨论了 Galton 有关遗传的工作，介绍了他的平均来说孩子在其父母亲和平均数的中间的发现。1918 年，在 Mendel 思想的基础上 Fisher 提出了一个机会模型[8]，它解释了 Galton 有关回归的发现，也解释了象身高那样的许多生物特征的近似正态性(第 5 章)。在付出引进许多复杂情况的代价之后，模型可以建立得非常逼真。本节以一种层层剥去的描述开始，必定不逼真，但是比较容易理解。稍后，将提及一些可能的精练。模型将集中于身高，尽管确定地对于其它的特征可以作相同的讨论。

模型中第一个假设是：

(1) 身高由一基因对所控制。

第二个假设是：

(2) 控制身高的基因以完全可加的方式起作用。

符号 h^*，h^{**}，h'，h''将被用来表示身高基因的 4 个典型变量。于是假设(2)意即，举例来说，h^* 对于一个人的身高总是贡献一确定的量，不管它是与 h^* 结合，还是与 h' 结合，或者与随便什么其它相结合。这些基因发挥作用非常不同于在 Mendel 豌豆中控制种子颜色的那些 y 和 g：当 g 与另外一个 g 结合时，它对于种子颜色贡献

绿色，而当它与一个 y 结合在一起时就没有作用。身高基因所起的作用更多地象上面 496 页习题 2 和 3 中的金鱼草基因。

根据假设(2)，每个基因对于一个人的身高贡献一确定的量。这个贡献(设以英寸计)将用表示基因的同一字母来表示，不过是大写。于是，具有基因对 h/h' 的人将具有等于和 $H+H'$ 的身高。起初，字母代表基因，其次场合，字母代表他们对于身高的贡献。

依据 Mendel，Fisher 假定一个孩子从他的父母亲的每一位随机地获取控制其身高的基因对的一个基因，如图 3 所示。更精确地说，父亲有一个控制身高的基因对，母亲亦如此。于是随机地从父亲的基因对中抽取一个基因，从母亲的基因对中随机地抽取一个基因，以构成孩子的基因对。

为了便于论述，假设父亲有一个基因对 h^*/h^{**}，母亲有一个基因对 h'/h''。孩子有½机会从父亲那里得到 h^* 和½机会得到 h^{**}。因此，父亲对于孩子身高的期望贡献是 $\frac{1}{2}H^*+\frac{1}{2}H^{**}=\frac{1}{2}(H^*+H^{**})$ 即父亲身高的一半。这就是 Galton 回归系数½的来源。同样地，母亲的期望贡献等于她的身高的一半。换句话说，如果你取大量的孩子，他们的父亲身高确定在一个水平上，而且母亲的身高确定在另一个水平上，那末这些孩子的平均身高必定大约等于

(3)　　　　½(父亲的身高＋母亲的身高)

表示式(3)称为父母亲身高的中间数。例如，就许多家庭来说，其父亲高 72 吋和母亲身高 68 吋，父母身高的中间数是½(72 吋＋68 吋)＝70 吋，平均说来孩子在成熟时将大约高 70 吋，加减一个小的机会误差。这就是 Galton 所发现统计规律性的生物学解释，回归到平常律。

假设(1)，即身高由一个基因对控制，在讨论中并不真正地需要，它的建立是为了避免复杂的和。譬如，如果牵涉三个基因对时，仅需对孩子假设遗传效应的可加性和从每一对中获取一个基因的随机性，如图 4 所示。

图 3 关于身高的遗传确定的简化了的 Mendel—Fisher 模型。身高由一个基因对控制，具有完全可加的遗传效应。从父母亲中的每一位的基因对中随机地获取一个基因构成孩子的基因对。

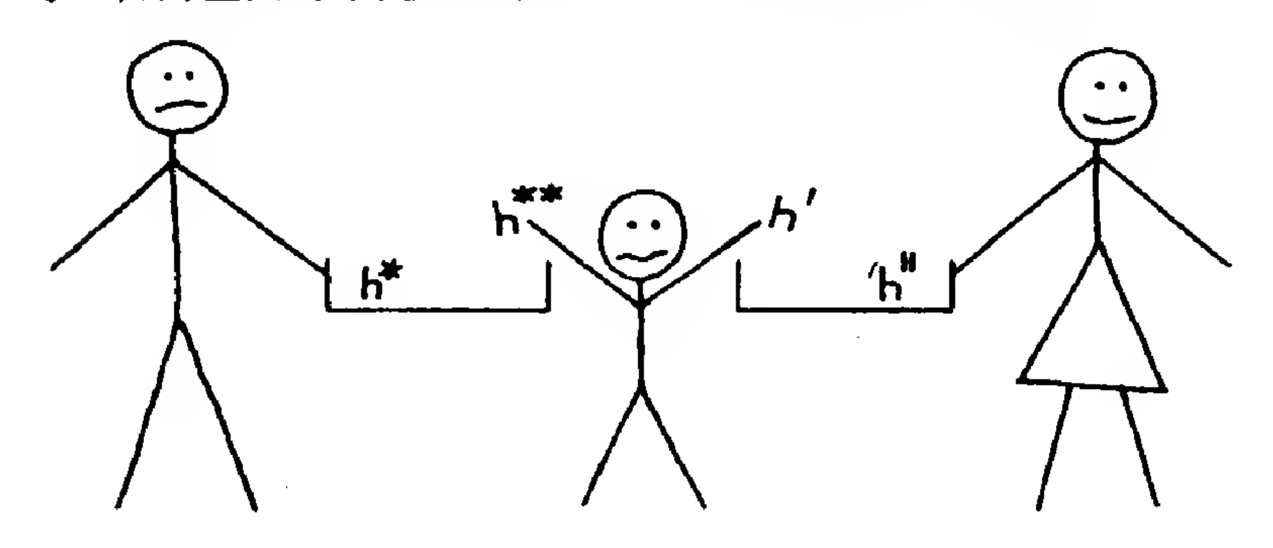
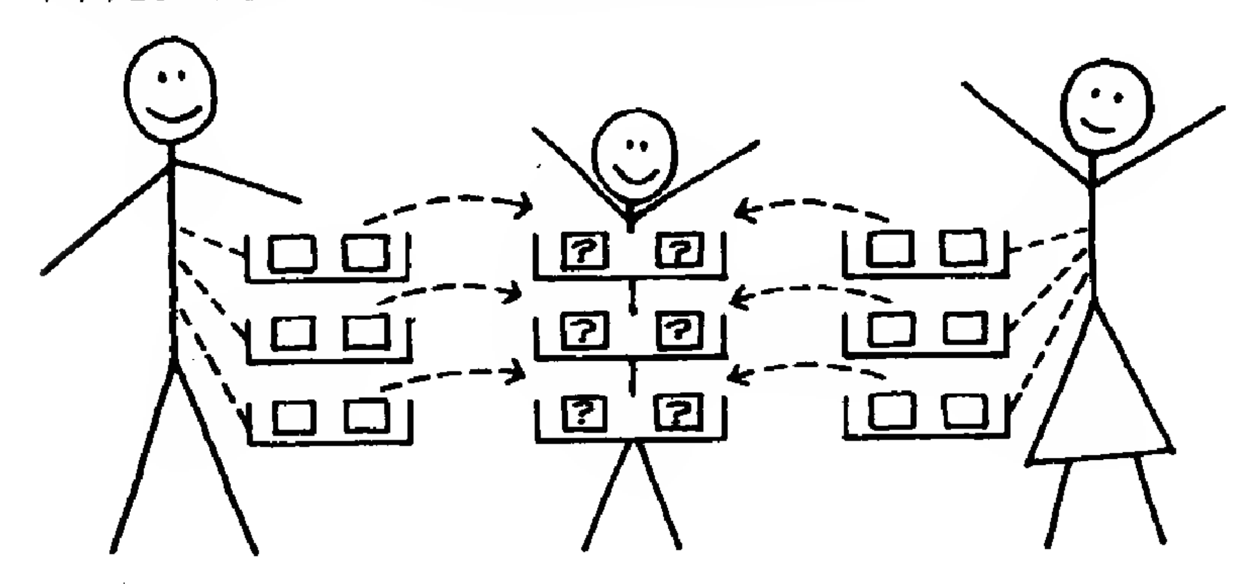


图 4 假设具有完全可加效应的三个基因对下，关于身高的遗传确定的简化了的 Mendel—Fisher 模型。从父母亲中的每一位的基因对中随机地获取一个基因组成孩子相应的基因对。



迄今为止，模型没有考虑身高中的性别差异。致力于此的一个办法是通过"调整"妇女的身高，使她们增加大约 8%从而恰好与男人一样高。也可以得到更漂亮的(及更复杂的)方法。

模型拟合有多好呢？根据在第三部分中讨论的 Pearson-Lee 研究，儿子的身高关于父母的身高的回归近似地为[8]

(4)儿子身高的估计

$$=15\text{ 吋}+0.8\times\frac{\text{父亲的身高}+1.08\times\text{母亲的身高}}{2}$$

回归系数 0.8 明显地低于由完全可加遗传模型所预测的 1.0。有些差异可能归因于环境效应，有些则归因于不可加遗传效应。还应该记得儿子平均比父亲高 1 吋。这也不能用完全可加遗传模型来解释。[10]（在该研究中有 1 078 个家庭，因此在平均数中 1 吋的机会变差是相当靠不住的。）

父母亲身高之间的相关系数大约是 0.25；因此儿子的身高关于父亲的身高的回归非常接近于

(5) 儿子身高的估计＝35 吋＋0.5×父亲的身高。

现在通过假设在双亲的身高之间不存在相互关系，从可加模型中的等式(3)也可以得出等式(5)[11]。然而，这基本上是两个错误抵消了的情况。可加模型有点不对头，以及双亲的身高有些相关。但是这两个事实在相反方向起作用而在等式(5)中相互抵消。

技术性注。从模型导出等式(3)时，没有必要对从不同的基因对中抽取的独立性作出假设；要紧的是每个基因有 50％的机会被抽取。没有必要对不同的父母亲中的基因之间的统计关联性（诸如独立性）作出假设。也没有必要对总体中基因的分布（譬如是均衡的）作出假设。

4. 模型的评价

遗传学描述了统计方法最令人满意的应用之一。回顾其发展，Mendel 发现了一些引人注目的经验规则——例如隐性状态在四分之一的第二代杂交中再现。他拼凑了一个含有现在称为基因的机会模型去解释这个规则。通过纯粹的推理，他发现了这些基因——他从未见过任一种。Galton 和 Pearson 独立地发现了另外一个引人注目的经验规则；平均来说，儿子的身高处在他的父亲和所有儿子的平均身高之间的半道处。

乍一看，Galton-Pearson 的结果看来很不同于 Mendel 的结果，很难看出他们两者如何可以用相同的生物学机制加以解释。可是 Fisher 能够做到。他解释了为什么孩子们的平均身高等于父母身高的中间数，甚至解释了为什么与平均数存在偏差。这些偏差是

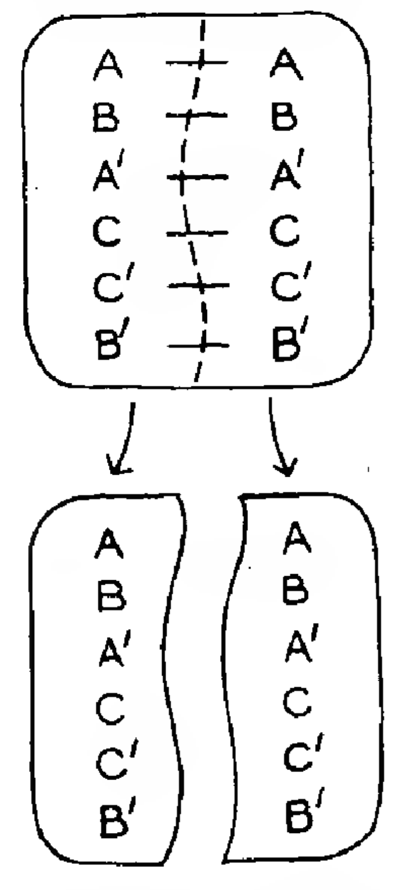
由于基因随机地从父母那里选取而传给孩子时的机会变异引起的。

机会模型现在用于许多领域。可是通常它们仅仅断言某些实体表现成如同它们是通过从一个盒子中随机地抽取一张票来确定的，而对于随机性的要求几乎不花什么努力去建立其物质基础。这些模型很少明确说明什么象盒子，或者什么象票。遗传模型是十分与众不同的，在这个模型里回答了这些问题。

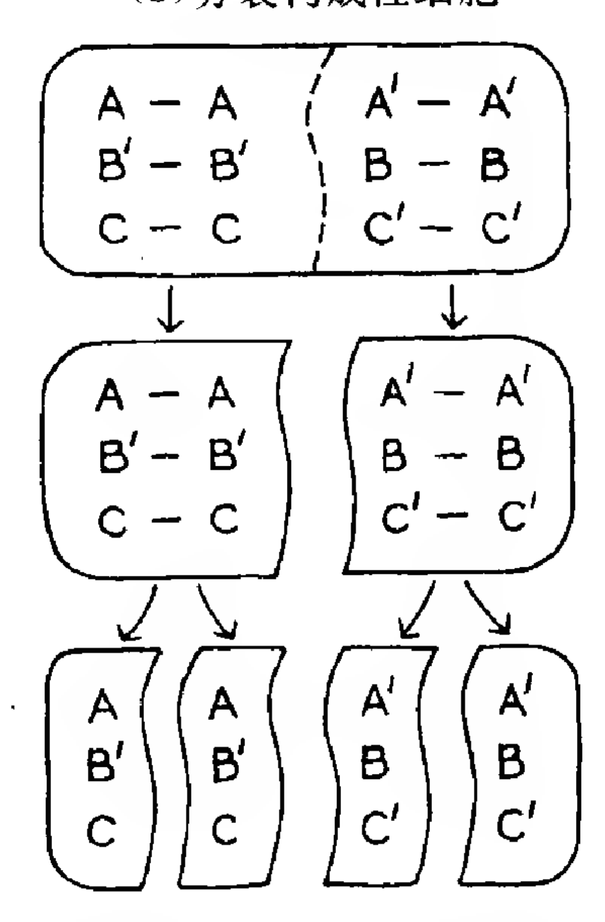
简短地，在模型中有两个主要的随机性来源：一为染色体（每一对中抽出一个）对性细胞的随机分配；另一个是性细胞的随机配对以产生受精卵。这两个随机性来源现在将更详细地讨论。

图5 通过分裂产生性细胞和躯干细胞。在这里染色体用大写字母表示，如C。染色体成为同源对，用撇表示。于是，C与C′形成一同源对；它们在化学上类同，但不恒同。在分裂的开始，一个细胞使所有它的染色体增加一倍。当对折时，C将表示为C—C。这两片在化学上是恒同的，且松散地贴在一起。同样地，C′对折成为C′—C′。

(a)分裂构成躯干细胞



(b)分裂构成性细胞



染色体合乎自然规律地成为同源的对。与C配伍的那个用C′表示；染色体C与C′在化学上类同，但不恒同。一个基因对有一个

基因位于同源染色体对的每一个染色体上。一个躯干细胞能够剖开形成其它细胞。在开始阶段，在亲本细胞中的每一个染色体自身对折，如图2和图5所示（图解地）。当染色体C处于该对折条件时，它将被表示为C—C。这两片在化学上恒同并且松散地连在一起。普通的躯干细胞的产生在图5a中给出。亲本细胞分裂为两个。每一个裂片变成一个单独的细胞，它带有每个对折了的染色体中的一半，外面完全地用如同（对折之前的）亲本细胞一样的染色体补足物包裹着，关于所产生的染色体不存在什么随机性——它是完整集合的复本。同源染色体不以任何特殊的方式处理。

性细胞的产生在图5b中给出。对折的染色体进入一个位置，这个位置具有来自每一个同源对的一个对折的染色体，这些同源对处在细胞将分裂的线的另一边，线的哪一边呢？这似乎是随机的，就象扔硬币。有时候一边，有时候另一边，恰如一枚硬币有时候正面着地，有时候背面着地。在模型中，它被假定是随机的。这第一步的结果在图5b的顶端给出。接下来，细胞分裂如图5b中间所示。每一片裂片包含对折了的染色体——但是每一个同源对中只有一个染色体呈现。最后，这些裂片的每一个再分裂，如图5b中底端所示，第二次分裂的结果是性细胞。[12]图5b的顶端同源对的集中是关键性一步。其结果是性细胞包含了普通的，未对折的染色体——但是每一个同源对中只出一个染色体。哪一个？随机地选择一个，这就是孟德尔学派遗传学中随机性的一个确切来源。

从产生的许多性细胞中的一个雄性细胞与一个雌性细胞的结合形成了一个受精卵。哪两个雄雌性细胞？这似乎是随机的，就好象从一个盒子中随机地抽取票。在模型中，它被假定是随机的。这就是孟德尔学派遗传学中随机性的第二个主要确切来源。

当考虑任意其他的机会模型时，问两个问题是有益的：

- 假定起票与盒子作用的确切实体是什么？
- 它们真是那样起作用吗？

5. 复习题

1. Mendel 发现了对于豌豆，未成熟的豆荚是绿色或黄色的。它们的颜色由一个基因对控制，变异体 g 代表绿色，变异体 y 代表黄色，g 是显性状态的。在一组培育试验中，已知豆荚颜色但不知遗传结构的植物互相杂交。结果列表如下。对于表中的每一行，猜测每一种亲本的遗传结构——g/g，g/y，或 y/g，y/y[13]。

亲本的豆荚颜色	具绿色豆荚的后代数	具黄色豆荚的后代数
绿×黄	82	78
绿×绿	118	39
黄×黄	0	50
绿×黄	74	0
绿×绿	90	0

2. Mendel 发现豌豆种子要么是光滑的要么有皱纹。他培育了一个纯粹光滑的品系和一个纯粹有皱纹的品系。使这两个品种杂交产生了第一代杂交，全都为光滑的。Mendel 使第一代杂交自身进行杂交的得到第二代杂交；在 7 324 颗第二代杂交中，有 5 474 为光滑的和 1 850 为有皱纹的。构造一个遗传模型去说明这些结果。在模型中，在光滑种子的期望频数与观察频数之间的一致如同 Mendel 所报告的那样接近的机会是多少？

3. 豌豆在三个不同的时期开花：早期，中期，和晚期。[14]培育试验给出如下结果：

早期×早期→早期

早期×晚期→中期

晚期×晚期→晚期

你从杂交

中期×中期

预期得到什么结果？如果你有 2 500 株来自中期×中期杂交所

产生的作物，1 300 株或以上为中期开花的机会是多少？

4. 在人类中，存在一个确定性别的专门的染色体对。男性有 X－Y 对，而女性有 X－X 对。一个孩子从母亲那里自动地得到一个 X 染色体；从父亲那里，有一半机会得到一个 X 染色体从而是女的，有一半机会得到 Y 从而是男性。某些基因只能由 X 染色体而转续；它们被称作性连锁的。一个例子是男性无毛的基因。（色盲和血友病是其他的性连锁特征。）

   (a) 如果一名男子有一个秃头的父亲，他较可能成为秃头吗？

   (b) 如果一个男子的外祖父是秃头，他较可能成为秃头吗？

   简短地解释。

5. 镰形血球贫血症是人类的一种遗传性疾病。在美国，特别在黑人中间流行，他们那里每 400 人中有 1 人忍受此痛苦。该病是由一个基因对控制，变量为 A 与 a，其中 a 引起疾病但却是隐性状态的：

   A/A，A/a，a/A——健康人

   a/a——镰形血球贫血症。

   (a) 假定双亲之一有基因对 A/A。孩子可能有镰型血球贫血症吗？

   (b) 假定父母亲都没有镰形血球贫血症，孩子可能有吗？

   (c) 假定父母都有镰形血球贫血症，孩子能避免患此病吗？

6. 小结

1. 只要繁殖是有性的，遗传机制是建立在基因对上的。后代获得的每一个基因对中的一个基因是从母亲生物体相应的基因对中随机地抽取，另一个是从父亲生物体相应的基因对中随机地抽取。一对中的两个基因是非常类同的，但不是恒同的。

2. 基因对可以以一些方式控制生物特征。一种方式是显性的。在这种情况，可能只存在两个品种（等位基因）的基因，设为 d 和 r。基因对 d/d，d/r，r/d 全都产生显性状态特征，而 r/r 产生隐性状

态特征。(豌豆的种子颜色是一个例子。)另一种方式是可加性。在这种情况，每一种基因有一个效应，基因对的效应是两个效应之和。(金鱼草的花色是一个例子)。

3. Fisher 证实了 Galton 的回归律是在只假定可加效应下的遗传机制的数学推论。

4. 遗传模型(至少部分地)解释了为什么孩子象他们的父母亲，也解释了为什么有差异。

第八部分　显著性检验



26

显著性检验

谁不愿说掩饰[对法律的注释]增加怀疑和无知？理解注释是比理解事物更为重要的事。

——Michel De Montaigne(法国　1533—1592)①

1. 引言

是由于碰巧，还是由于其他?统计学家发明了显著性检验来处理这类问题。当今，读一篇科研文章而不遇到检验和显著性水平是几乎不可能的。因此，弄清他们的意义是有益的。第 26 章至第 28 章的目的是要解释显著性检验背后的含意和所用的语言。某些限制将在第 29 章中指出。这一节先举一个例子。

假如一位参议员提出一条简化税法的议案。这位参议员声称这项简化不会改变税收总数，即年终结算税收入不会改变。用微型模拟模型来估算这类改变对政府政策的影响，这时显著性检验就起了作用，虽然细节是繁复的，思想却是简单的。

为了估算参议员的提案要求，财政部应用了计算机档案中 10 万张有代表性的完税申报表，每张表列出在旧税则下应付的税款总数。从表中详细的记载，财政部能算出在新税则下的总税收，然

后再看一下二者之差。

新税则下的税收数－旧税则下的税收数

这个差是正号还是负号十分关键。正号意味着在新税则下，财政部将从纳税收人那里收到较多的税；负号呦，他们就将收得少一些。参议员认为平均差数应为零——多收和少收恰好相抵。

在我们(部分假定)的例子中，我们从档案中随机地选择 100 张完税申报表。所构成的一个试点样本[2]，样本差的平均数为－219 美元，因此看来新税则要使财政部减少一些收入，但标准差却很大，725 美元。下面是一位国会助理和一位财政部官员对计算结果所做的讨论。

助理： 我先谈，我不相信 SD。怎么它会比平均数大这么多？

财政部： 嗯，申报完税遍及全国各地。有些人一分不付，那有大约 20％的申报，就在那里。有些人付几千元，直至有付几十万的。这些数字有一条长尾巴[3]。新税则对某些纳税人确实显示很大的差异，而对另外一些人却没有差异，我们用齿很细的梳子梳理这一程序，请相信我，它们是对的。

助理： 现在我想你将告诉我我们的建议毕竟不是居中不变的。

财政部： 如果议案被通过，对每一张申报表，我们少收 200 美元。你可能不认为它是什么严重的事，但全美国有一亿张申报表。因此我们是在讲二百亿美元或更多一点。

助理： 请等一等。你只是用一个 100 张申报表的样本来得出这样的结论，对不对呀？

财政部： 对。

助理： 而你在不断地告诉我 SD 是 725 美元。因此这个 219 元的差数只是一个 SD 中的一小部分。如果今后我有幸见到的话，那是一个机会变异。

财政部： 不，不。我们需要的是 SE，不是 SD。可计算 SE，我们应该有一个盒子模型。盒中有 100 000 张票，每一张代表

档案中一张申报表，票上的数字表示新税则下的税款和老税则下的税款之差。我们随机地抽取 100 张票作为我们的样本。数据就象这一百次抽得的数。对吗？

助理：　不错，但它们告诉我们什么呢？

财政部：　我们所争论的是盒中 100 000 张票的平均数——新税和老税的平均差。你说它是零元，我们说它是负数。

助理：　对。我说，由于抽取的运气，你在你的样本中得到较多个较大的负数，请记住 SD。

财政部：　嗯，那是下一个问题，我们不知道盒子的 SD，我们可以用样本的 SD 来估计它为 725 美元。

助理：　那看来是合理的，下一步将向那里走？

财政部：　由盒中抽得 100 个数之和的 SE 是 $\sqrt{100}\times 725=7\ 250$ 美元，它可能就是全体 100 张样本申报表上数字之和的机会误差。我们正在考虑平均数，所以要除 100 而得 $7\ 250\div 100=72$ 美元。

助理：　然后怎样？

财政部：　好罢，暂时假定你是对的，即盒子的平均数确是零美元。于是，你必定期望样本的平均数在零的附近。而我们所得到的是 -219 美元，它是在你的期望值之下 3 个 SE：

$$\frac{-219-0}{72}\approx -3$$

助理：　嘿…

财政部：　我们这里已抽得足够的数，因而可以应用正态近似。正态曲线下，-3 左边的面积不到千分之一，你在说的是 1 000 次中不到 1 次的机会。

助理：　可能是，但哪里出来个正态曲线？数据的直方图肯定不象正态。

财政部：　对，我们是对抽得数的平均数的概率直方图而应用正态曲线的。

助理：　同意，我理解你现在在做什么。

财政部：　现在你可以如参议员所希望的那样，仍然认为盒子的平均为零美元，或者和我们一样，认为它是负的。但是如果你坚持零美元，你需要一个小小的奇迹来帮助你解释为什么样本平均在 1 000 次中仅有一次机会在零美元之下那么远。

助理：　我想那几乎是一个矛盾，可能我将放弃它，你想此新税法将起多大的影响？

财政部：　新税法将使我们对每一申报表损失 200 元，它不是一个巨大的差异，但它是现实的，我意思是说你不能将它看作机会波动而置之不理。

我们与检验的第一次遭遇现在完成了。例子中所争论的那个问题一次又一次地出现：一方面认为差异是真实的，但怀疑者可能说它完全是机会波动，怀疑者将被一个机会计算挡开，如在以上的对话中那样，这个计算称为显著性检验，下二节将提出一个检验中的关键成分[4]。

习题 A

1. 在本节的例题中：

 (a)Jones 先生按新税法将付 3 292 美元，而按老税法付 3 117 美元。他的那张票将记上______。请填空并作简短解释。

 (b)Smith 夫人的票券记着 −753 美元，她的处境是在新的税法下好呢还是在老的税法下好？简短地解释。

2. 在本节例题中：

 (a)财政部的盒子模型有____张票券在盒子中，并抽了____次。选择：

 100　1 000　10 000　100 000

 (b)盒子的 SD 是____725 美元。选择：

 已知为　　由数据估计为

3. 在本节的例题中，假如抽样取得的 100 张完税申报单，其平均申报数按新税法为 3 182 美元，按老税法为 3 217 美元，而此 100 个差数的 SD 仍为

725 美元。现在谁赢：是财政部的官员还是国会的助理？

4. 一颗骰子掷了 100 次。总点数为 368 而不是所期望的 350。这一情况能解释为机会变异，还是骰子是灌过铅的？

5. 一颗骰子掷了 1 000 次，总点数为 3 680 而不是所期望的 3 500。这一情况能解释为机会变异，还是骰子是灌过铅的？

这些习题的答案在第 721 页。

2. 原假设和备择假设

在上一节的例题中有 100 纳税人的样本数据。双方都见到数据中 219 美元的平均差；用统计速记的讲法，我们称"观察到"219 美元。争论是在理解。财政部官员声称观察到的差异是"真实的"。这听起来可能有些奇怪；当然 219 美元是不同于 0 美元。但问题是，219 元是否如那位助理所说，正好反映机会变异；还是，是否它表示一个真实的改变——在新法则下税收将下降。这样的争论是如此地普通使得人们对此两种情况都曾给了专用名词。

- 原假设是指一个观察到的差异只反映机会变异。这就是上节对话中，那位国会助理的情况。
- 备择假设是指观察到的差异是真实的。它就是财政部官员的情况。

为了说服助理放弃原假设，财政部官员为此问题建立一个盒子模型，把原假设和备择假设翻译成有关模型的陈述。

- 原假设——盒子的平均数等于 0 美元
- 备择假设——盒子的平均数少于 0 美元

做一次显著性检验，原假设应表为盒子模型的一个陈述。备择假设通常也是如此。

这些术语可能使人迷惑不解；"备择假设"常是某些人所要证明的东西。这时，"原假设"是用机会变异的语言对所发现的东西作出的一种备择（并且笨拙的）解释。然而，这些名词已经用惯了，实在无法改进。

在上面的对话中，重要的是盒中的平均数，争论的是关于盒中 100 000 个数而不是样本中的 100 个数。一个显著性检验只能在有关盒子的辩论才有意义。（这一点将再在在第 29 章，第 5 节中讨论。）

每一合理的显著性检验总涉及一个盒子模型。检验是要查明一个观察到的差异是否为真，还只是一个机会变异的问题。一个真的差异确实说明盒子的某些情况，而不是仅仅抽样的一次侥幸成功或失败的反映。

习题 B

1. 是或非：备择假设说，除了机会变异，什么也没有进行。
2. 是或非：为了检验一个原假设，必须把这一假设表为有关一个盒子模型的陈述。
3. 一个显著性检验在有关____辩论中有意义。试在下列二条中选择一条，并作简短解释：

   数据　　　数据模型
4. 从一个盒子中，随机放回地抽取一百个数。这一百个数的平均是 102.7，它们的 SD 是 10，有人声称盒子的平均数等于 100。可信吗？如果所抽得的平均数是 101.1，那又怎样呢？

这些习题的答案在第 721 页上。

3. 检验统计量和显著性水平

在第一节的对话中，财政部官员对数据建立了一个盒子模型。为了便于讨论，他暂时假定原假设是正确的，因此盒子的平均数是 0 美元。在这一基础上，他计算出样本平均数的观察值与它的期望值相差有多少个 SE：

$$\frac{-219\text{ 美元}-0\text{ 美元}}{72\text{ 美元}}\approx-3$$

这是检验统计量的一个例子

检验统计量是用来度量数据与原假设下所期望的这二者之间的差距。

我们的检验统计量常称为 Z

$$Z=\frac{\text{观察值}-\text{期望值}}{\text{SE}}$$

并且所有应用这一个比的检验称为Z 检验。

请记住如下解释:Z 表明一个观察值与它的期望值相差有多少个SE。期望值总是从原假设出发计算出来的。在那个对话中,差的期望值是 0 美元。另外的原假设将给出不同的期望值。正是这个原假设告诉了我们在 Z 的分子中采用 0 美元而不是其它数字为水准基点。这恰好是原假设进入本程序之处。

在原假设的基础上,Z－统计量把观察值转换成标准单位

－3 的 Z 统计量阻止了助理的想法。为何它的威胁性如此强烈? 3 毕竟不是一个很大的数。理由当然是,正态曲线下,－3 左边的面积令人可笑地小。要取得一个样本平均数低于它的期望值三倍或三倍以上 SE,其机会大约是千次中一次。



(查表,面积是 1%的 0.135,四舍五入,我们得到 1%的 0.1。这就是 0.01 中的 0.1=0.001=1/1 000。)

1‰的机会压倒了助理,并迫使他承认新税则将降低总税收——不只是对一个样本而是对整个(美国)人民。这类的机会称为一个观察到的显著性水平,它是作为一个显著性检验的部份内容计算出来的。

为什么考虑－3 左边的面积呢？回答是，数据可能变动，从而 Z 也变动。例如，如果样本平均数是－239 美元，SD 是 590 美元，

$$Z=\frac{-239\text{ 美元}-0\text{ 美元}}{59\text{ 美元}}\approx -4.1$$

这是反对原假设的一个更有力的证据：对助理来说，0 美元之下 4.1 个 SE 比 3 个 SE 更坏。另一方面，若样本平均数是－162 美元，而 SD 为 630 美元，

$$Z=\frac{-162\text{ 美元}-0\text{ 美元}}{63\text{ 美元}}\approx -2.6$$

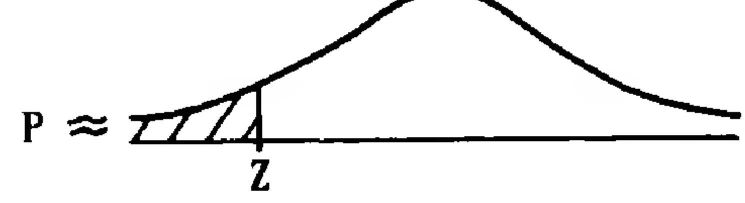
这是一个较弱的证据。－3 左边的面积代表那些会给出比观察值更极端 Z 值的样本——因而是反对原假设的更有力的证据。

> 观察到的显著水平是得到一个与观察值同样极端或更极端检验统计量的机会。机会是在原假设为真的基础上计算出来的。机会越小，反对原假设的证据越强。

观察到的显著水平常记作 P，表示概率，并常称为检验的 P－值。在例题中 P 大约是 1 000 分之 1。

Z 检验可以总结如下：

$$Z=\frac{\text{观察值}-\text{期望值}}{\text{SE}}$$



由于检验统计量 Z 依赖于数据，因而 P 也是如此。这就是为什么 P 被称为一个“观察”到的显著水平。

当正态近似可以用于抽得的平均数的概率直方图时，只要样本合理地大，我们就可应用 Z 检验。（平均数通过 Z 已被转换成标准单位。）对于小样本，统计学家喜欢采用 t 检验，我们将在下面第 6 节中讨论它。

到此，我们可以更清楚地看到Z检验的逻辑性。这是如此设计的一种否定论证，如果原假设为真，则将导致一个荒谬的结论，因而必须拒绝原假设。你着眼于数据，计算检验统计量和观察到的显著水平。比如，取P为1/1 000。为了理解这个数的意义，请先假定原假设为真。接下来，请设想许多其他的调研人员在重复你的实验。一千中之一的意思是你的检验统计量实在走得太远：一千个调研人员中仅有一个将得到一个检验统计量比你所得到的那一个一样地极端或更极端一些。原假设就是这样让你陷入矛盾之中，因此，你必须拒绝它。一般地说，观察到的显著水平愈小，你愈要拒绝原假设。在使用显著性检验拒绝原假设时，就是着重运用矛盾的反证论点。

运用矛盾的论证是复杂的，并且想象P给出原假设为真的机会是很富有诱惑力的，但千万不要被引入歧途，不论你曾抽过多少次，有关盒子平均数的原假设要么总是为真，要么总是错的。按照频率论，这里没有机会的问题在内（有关置信区间的一个相类似的问题在第21章，第3节内讨论）。而观察到的显著水平所给出的是获得否定原假设比手边证据一样强或更强证据的机会。

一个检验的P值并未给出原假设为真的机会。事实上，P是在原假设的基础上计算出来的。

习题C

1. (a)设其他事物都相同，问下列诸P值中哪一个对原假设最好，并作简短解释。

   1%中0.1，　3%，　32%，　77%

   (b)对备择假设重复上述的问题。

2. 是或非，并解释之：

   (a)观察到的显著水平依赖于数据

   (b)设观察到的显著水平为1%，那么100次中仅1次的机会原假设为真。

3. 按照某调研者的模型，数据就象从一大盒子中随机地抽取 400 次。原假设声称盒子的平均数等于 50，而备择假设声称盒子平均数大于 50。事实上，数据的平均数为 52.7，且 SD 为 25。计算 Z 和 P。你得出什么结论？
4. 在八十年代中，许多公司试验“自选时间”工作制度，公司领导定出宽广时间限度使雇员们在此限度内能自由选定适合他们的工作时间。在许多其它事物中，据说，自选时间能降低缺勤率。假定一家公司知道在过去若干年中，雇员们平均缺勤 6.3 天（不计休假）。今年，公司引入自选时间。领导选择了 1 个 100 简单随机样本的雇员进行仔细跟踪。一年下来，这些雇员平均缺勤 5.5 天，并且 SD 为 2.9 天。能否说自选时间减少缺勤率？还是这一样本只是一个机会变异？
5. 重复例 4，为把样本平均数及 SD 改为 5.9 和 2.9 天。

这些习题的答案在第 722 页。

4. 做一次显著性检验

做一次显著性检验是一件相当繁杂的工作。首先需要一些数据。其次，一位调研者应该

- 把原假设译成结合数据的盒子模型；
- 定义一个检验统计量以度量数据和原假设下期望值二者之差。
- 计算所观察到的显著水平 P。

检验统计量的选择依赖于模型和所考虑的假设。我们至今所讨论的是“单尾 Z 检验”，且基于 Z 统计量。（双尾检验将在第 29 章，第 2 节中讨论）。还有依据 t 统计量的“t”检验（见下面第 6 节），依据 χ^2 统计量的 χ^2 检验（第 28 章），和许多其他在本书中甚至没有提到的检验。然而，一切检验都按照上述步骤进行，并且它们的 P 值可以按同样的方式予以理解。

很自然地会问把观察到的显著水平定得多小使调研者可以据以拒绝原假设。许多统计工作者在 5% 处划条线。

- 若 P 是低于 5%，我们称此结果为统计显著。

另有一条线在 1%。

·若 P 是低于 1%，我们称此结果为高度显著。

这些分界线将再在第 29 章第 1 节中讨论。

不要让这些行话打扰你从而偏离主题。当数据与按照一种理论将作的预测相距太远的话，这种理论必须予以拒绝。在统计学中，当观察值距离期望值好几个 SE 的话，原假设一定要拒绝。

本节的练习将帮助你理解 P 值。建立盒子模型是显著性试验中关键的第一步；下一节的习题将教你怎样建立模型。你应该做这些习题。

习题 D

1. 在第一节的对话中，财政部官员运用 Z 检验拒绝____的平均数为 0 美元的原假设。用下列二选择之一

盒子　　　　样本

填入空格并作简短解释。

2. 是或非：

(a)一个"高度显著"的结果不可能是由于机会的缘故。

(b)如果一个差是"高度显著"，原假设为真的机会小于 1%。

3. 是或非：

(a)若 P 为 43%，则原假设看上去是可信的。

(b)若 P 为 1%的 0.43，则原假设看上去是不可信的。

4. 是或非：

(a)若观察到的显著水平为 4%，则此结果为"统计显著"。

(b)若一检验的 P 值为 1.1%，则此结果为"高度显著"。

(c)若一个差是"高度显著"，则 P 小于 1%。

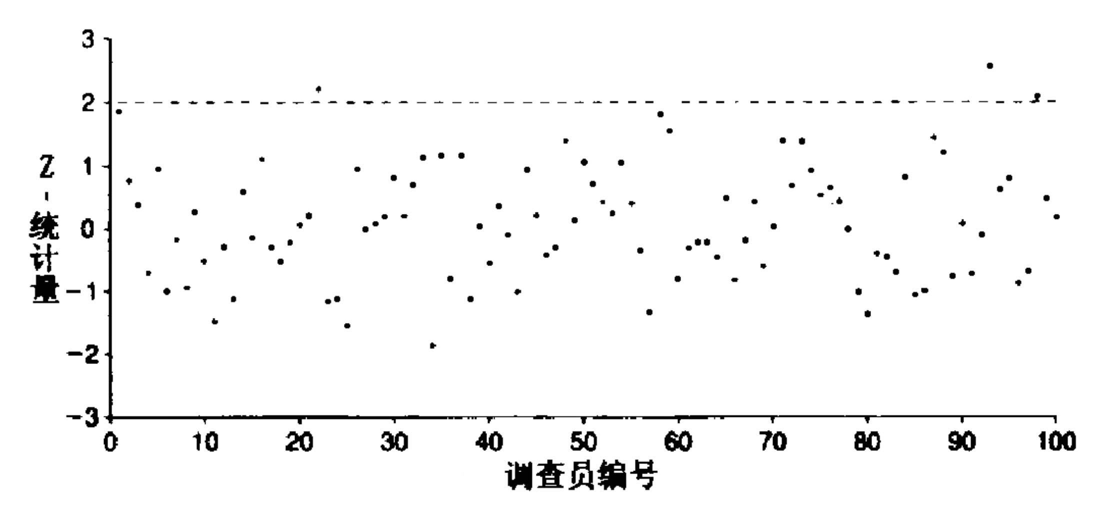
(d)若观察到的显著水平为 3.6%，则 P=3.6%。

(e)若 Z=2.3，则观察到的值高出于原假设下所期望的值 2.3 个 SE。

5. 一个调研人员从一个盒子中随机放回地抽取 250 张票。问抽取的平均数高出盒子平均数 2 个 SE 的机会是多少?

6. 一个盒子含有记数的票。一百个调研人员准备检验盒中平均数等于 50 的原假设。每个调研人员随机放回地抽取 250 张票，计算每人所抽的票的平均数并做 Z—检验。所得结果描绘于下页图：第一号调研人员得到的 Z 统

计量为 1.9，于是标绘在图上为点(1，1.9)；第二号调研员得 Z 统计量为 0.8，标绘在图上为(2，0.8)；余类推。调研人员并不知道，原假设恰好为真。



(a)是或非，并解释：Z 统计量为正数，当抽取的平均数大于 50 时。

(b)多少调研人员应取得正的 Z 统计量？

(c)多少调研人员应取得大于 2 的 Z 统计量？有几位确实取得这样大的 Z 统计量？

(d)若 Z＝2，P 是多少？

这些习题的答案在第 722 页上。

5. 0-1盒子

Z 检验也能用在涉及分类及计数的情况中。它只是把 0 或 1 放在盒子中(第 17 章第 5 节)。本节给出一个例子。此例有另一特点：备择假设不能译成有关盒子模型的陈述。

1973 年，Charles Tart(查尔斯・塔特)在加州大学 Davis(戴维斯)分校做了一个实验以验证超感觉力(extrasensory perception 简记 ESP)⑤。Tart 应用一台机器叫做“Aquarius”。它包含一个电子随机数发生器和四个“标记”。应用随机发生器，这台机器随机地选取四个标记中的一个，但它并不指出哪一个。然后参试对象猜一下哪一个标记被选中，并揿一下按钮。最后，如果参试对象所猜的标记是正确的话，机器亮起它所选中的标记的灯，并响起铃声，机器(自动)记录试验次数和正确猜测的次数。

Tart 从过去曾显示超人视力的人中选取 15 个参试对象。每一对象在 Aquarius 机器上猜 500 次，总共 15×500＝7 500 次猜

测，其中 2 006 次是正确的。而今即使对象没有任何程度的超视能力，他们仍有大约 1/4 的次数是正确的。换句话说，完全受机会的影响。大约有 $1/4\times7\ 500=1\ 875$ 次可以期望是正确的，确实，存在着

$$2\ 006-1\ 875=131$$

次正确的猜测这一余额，但能把这理解为机会变异吗？

Tart 能够（并实施了）通过做一次显著性检验以否定那个解释。为检验建立一个盒子模型，你假定 Aquarius 按照完全随机的程序产生数字，因而在每次试验中，四个标记中的每一个恰好在四次中有一次机会被选中。现在你假定不存生 ESP。于是每次猜测恰好有四分之一的机会是正确的。数据包含着 7 500 次猜测的记录，表明哪一次是正确的，哪一次是不正确的。原假设表明数据象从盒子

| 1 | 0 | 0 | 0 | 1＝正确，0＝不正确

中抽得的 7 500 个数。因此正确猜测的次数就象从盒中 7 500 次抽得数之和。Aquarius 机把每一次猜测分类为正确或不正确，并记下正确的猜测数。这就是为什么我们需要一个 0－1 盒子。这就完成了原假设的盒子模型。

没有一种切合实际的方式把备择假设转化入一盒子模型中。理由是：如果参试对象确有 ESP 的话，每次猜测为正确的机会可能在很大程度上依赖于以前的试验的结果，并且可能随一次次试验而改变。于是数据就不象从一个盒子中的抽取数[8]。

一旦原假设被转化入一盒子模型中，很容易应用

$$Z=\frac{\text{观察数}-\text{期望数}}{SE}$$

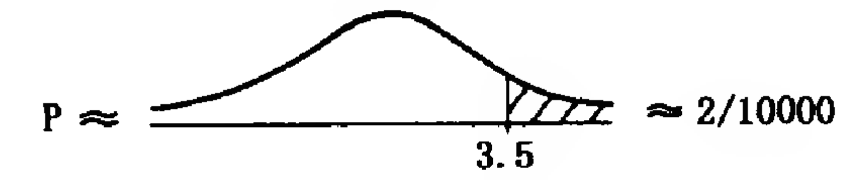
去检验。“观察数”2 006 为正确猜测的次数。正确猜测的期望数来自盒子模型，并等于 1 875。Z 统计量的分子等于 $2\ 006-1\ 875=131$ 是正确猜测的剩余数

现在来讨论分母。对于正确猜测数你需要 SE。在本例中，原假

设精确地告诉了你盒子中所包含的内容：一个“1”三个“0”。盒子的 SD 等于$\sqrt{0.25\times0.75}\approx0.43$。SE 等于$\sqrt{7\ 500}\times0.43\approx37$。因此

$$Z=\frac{131}{37}\approx3.5$$

观察值 2 006 在期望数之上 3.5 个 SE，因而 P 相当小：



这是推翻原假设的一个明确的证据。但又很难把正确猜测的余额解释为机会变异。当然，这并不能证明 ESP 的存在。比如，Aquarius 随机数发生器可能不怎么完好（第 29 章第 4 节）。或者机器会给参试对象某些难以捉摸的暗示如何挑选一个标记。这些结果除了 ESP 外可能有很多合理的解释。但机会变异却不是其中之一。这就是显著性检验所显示的，ESP 的例子完毕。

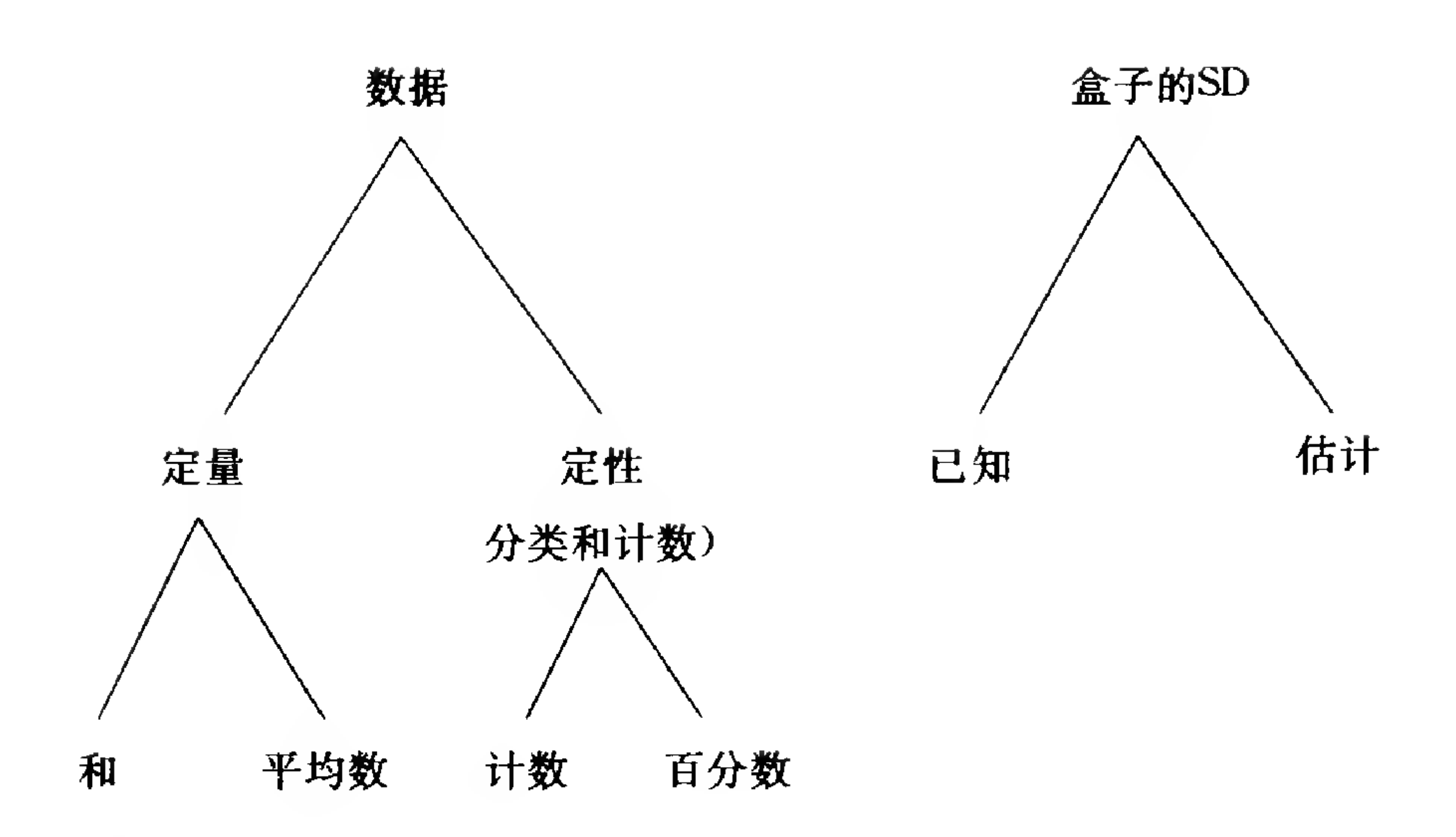
我们对 ESP 如同对税法的例子，使用相同的 Z 统计量

$$Z=\frac{\text{观察到的}-\text{期望的}}{\text{SE}}$$

但这里的 SE 是一个计数的 SE，而不是一个平均数的 SE。在你着手以前首先要决定你观察的是什么：你是在与什么打交道。一个和、一个平均数、一个计数、还是一个百分数？它会告诉你用什么公式

对定量数据，盒子的 SD 通常是不知道的。这就是税法例题的情况：财政部官员必须从数据中估计出盒子的 SD。然而，如果原假设给出盒子的 SD，你就不必估计他了。

当问题涉及的是（定性数据的）分类或计数时，需要一个 0-1 盒子模型。原假设常常指明盒子中“1”的个数所占的分数。你可能直接从原假设计算盒子的 SD 而不是从数据中将它估计出来。（见下页图）



习题 E

1. 在 Tart 的实验中，原假设声称______，试选下列诸选择之一，填入上面的空白。

(i)数据象从盒子 [0] [0] [0] [1] 中 7 500 次抽得数。

(ii)抽得数的平均数象盒子的平均数

(iii)盒子的平均数接近于抽得数的平均数

(iv)ESP 就象机会一样

2. 一枚硬币抛 10 000 次，头像朝上 5 167 次。是否呈头像的机会等于 50%？或者对此情况是否头像朝上的太多一些？

(a)试用盒子模型的术语，写出原假设和备择假设。

(b)计算 Z 和 P。

(c)你得出什么结论？

3. 重复习题 2，如果头像朝上出现 5 067 次，如同 Kerrich（见第 16 章第 1 节）所做的那样。

4. 作为一个统计科研项目的一部分，Frank Alpert 先生某一天询问在加州大学，贝克莱分校的 Sproul 广场上所遇到的首 100 个学生他们在哪一个学院注册上学。他的样本包括 53 名男生和 47 名女生。从注册主管者的数据，那一个学期，贝克莱分校共有 25 000 学生而 67% 为男生。问 Frank Alpert

的抽样程序象取一个简单随机抽样吗?

填空。它们将一步一步地引导你走向原假设的盒子模型。(对此盒子是没有备择假设的。)

(a)对每个______在盒子中有一张票。

样本中的人　　那学期在 Berkeley 注册的学生

(b)对男生票上标记______,对女生票上标记______。

(c)盒子中的票数是______而抽取次数是______

选择:100,25 000。

(d)原假设声称样本象从盒子中随机地抽取____ ____。(第一个空白必须填一数字,第二个空白填一言词。)

(e)盒子中"1"所占的百分数是______。选择:53%,67%。

(f)这是______________。

从数据中估计得来的　　由原假设给定的

5. 本题为习题 4 的继续。填空,这将逐步引导你到 Z 和 P。

(a)观察到男生人数是__________。

(b)男生期望人数是__________。

(c)该期望数是__________。

由原假设算得　　由数据估计得来

(d)如果原假设为真,样本中的男生数象盒子中抽得数之__________。选择:和,平均数。。

(e)男生人数的 SE 是__________。

(f)该 SE 是__________。

由原假设算得　　由数据估计得来

(g)Z=__________和 P=__________。

6. 本题为习题 4 和 5 的继续。问 Alpert 的抽样程序象不象取一个简单随机抽样? 回答是或否,并简短地解释之。

7. 另一个 ESP 实验应用了一台"十个选择器"。它象 Aquarius 机,只是具有 10 个标记而不是 4 个。假定在 1 000 试验中,一个参试对象得到 173 次正确猜测数。

(a)建立一个盒子模型的原假设。

(b)该盒子的 SD 是______。用下述选择填空,并简短地解释之。

$\sqrt{0.1\times 0.9}$　　$\sqrt{0.173\times 0.827}$

(c)作 Z 检验。

(d)你得出什么结论?

8. 一个实验室小鼠群体共数百只,它们平均体重约 30g,SD 约 5g。作为一个实验的一部份,研究生们被通知不用任何确定的方法任意选取 25 只小鼠⑦。这些动物的平均体重约 33g,SD 约 7g。任意选取是否与随机抽样一样?或者说 33g 比该群体的平均体重高出许多。简短地讨论之,用盒子模型的术语提出原假设,并计算 Z 和 P。(没有必要提出盒子的备择假设)

9. (难)折扣商店常以特低价向顾客推销新商品诱使顾客试用。然而一位心理学家预测这种做法只会降低营业额。在一组折扣联营商店合作下进行一个实验以检验此预测⑧。选择 25 对商店并按照诸如店址及销售额等特征进行匹配。这些商店不做广告并用同样的方式布置商品。在全体 50 家商店中引入一种新的饼干。在每一对店里随机地选取一家用特别低的价格销售这种饼干,二星期后把价格回复到正常水平;在这对店的另一家里始终按正常价格销售饼干。从销售饼干开始日起六个星期后结算每一商店的总销售额。在 25 对的商店中有 18 对,那些一开始就按正常价格出售饼干的店结算下来却比同组另一商店销售较多的饼干。问这种结果能解释为机会变异吗?还是它支持了预测,即以低价格推销商品会减少长期销售?(用盒子模型的语言提出原假设;这里没有有关盒子模型的备择假设。)

这些习题的答案在第 723—724 页上。

6. t-检验

对小样本,Z 检验必须修改。统计学家们采用 t 检验,它是英国的 W·S·Gossett(1876-1936)所发明。他在牛津大学读得学位后进入几纳斯酿酒厂(Guinness Brewery)工作,并晋升该厂行政领导。他用笔名"学生"发表他的著作,因为他的雇主不愿意同行们知道这些结果可能多么有用。本节将用例题来阐明怎样做 t 检验。但是,由于讨论有点儿技术性,初读可以跳过。

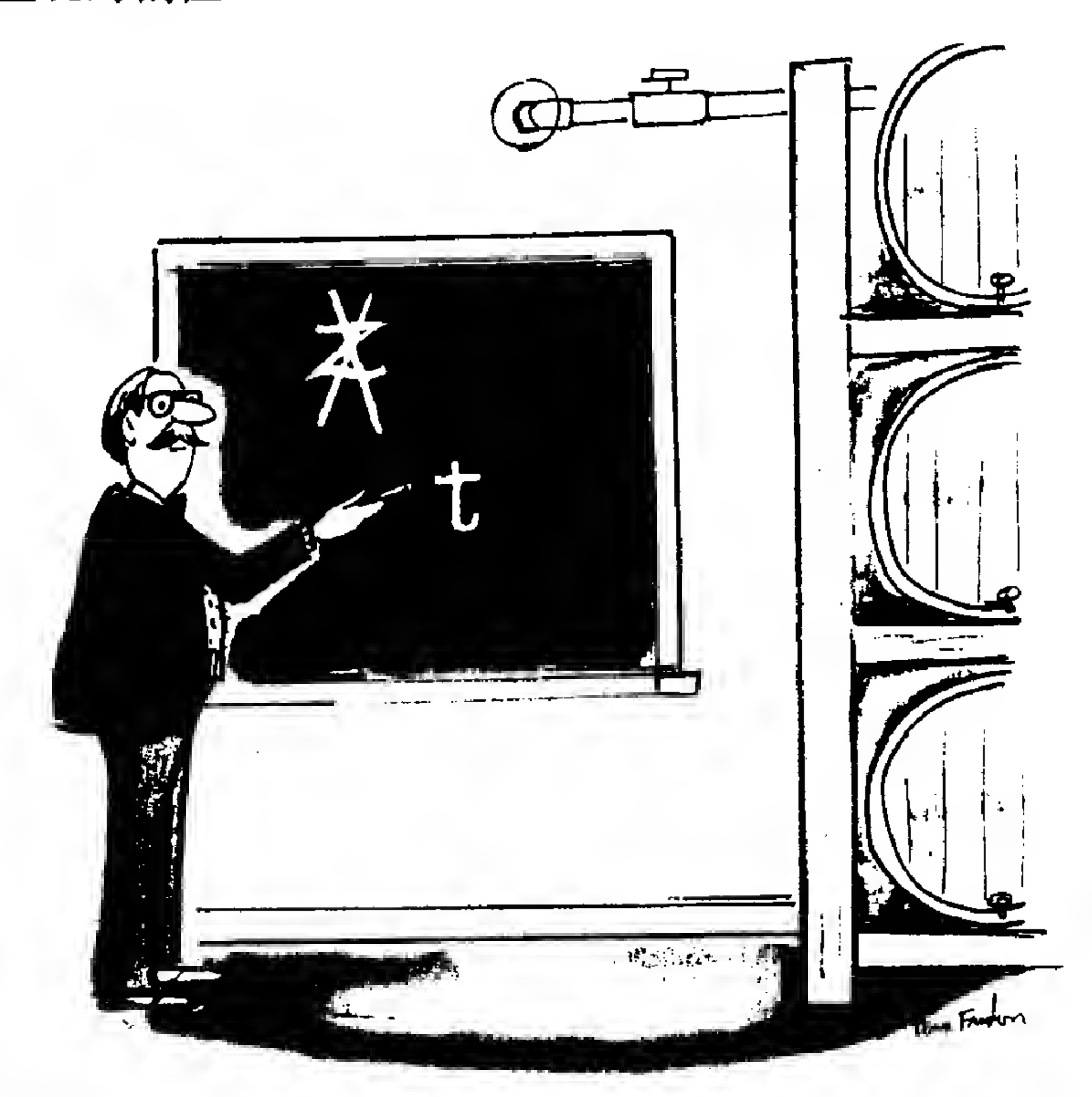
在洛杉矶,人们在具有各种交通流动情况的超高速公路附近进行很多一氧化碳的浓度的测定工作,其基本方法涉及用特殊的袋子捕捉样本空气,然后用一种叫做分光光度器(Spectrophometer)来测定袋样本中 CO 的浓度。这类仪器可度量浓度

高达 100ppm(按体积算的百万分率),误差为 10ppm 级。然而,这种仪器很灵敏,且必须每日校对。这一步骤要求用一台分光光度器去度量一个预先制好的气体样本叫做量标气体(span gas),其 CO 浓度精确地控制在 70ppm。如果分光器在量标气体上测出的读数接近于 70ppm,这台光度器已经就绪可以使用了,否则必须重新调整。一个复杂的因素是度量误差的大小逐日在变。在任何指定的一天,我们假定误差独立并遵循正态分布,具有依赖于日子的某未知 SD[⑨]。

某日,一位技术员对量标气体作了 5 次读数并得到

78　　83　　68　　72　　88

这 5 个数中有 4 个高于 70,且其中有几个大了很多。这一点能否在机会变异的基础上给以解释?还是也许由于光度器不正常的调整而呈现的偏性?



我们需要一个显著性检验，且为此还需要一个盒子模型。所要用到的是高斯模型，(第 24 章第 3 节)。根据这一模型，每一测定数等于 70ppm 的真值加上偏性，加上从误差盒子中有放回地抽取的数。误差盒子中的票平均为零，但它们的 SD 未知。

测定数 = 70 ppm + 偏性 + □ 误差盒

这里关键的参数为偏性。原假设是偏性等于零。在这一假设下，5 个测定数的平均有一期望值 70ppm；平均数和 70ppm 之差可解释为机会变异。备择假设声称偏性不等于零，因此测定值与 70ppm 之差是实在的。

如前，合适使用的检验统计量为

$$\frac{\text{观察值}-\text{期望值}}{\text{SE}}$$

五个测量值的平均数是 77.8ppm，而他们的 SD 是 7.22ppm。看来，误差盒子的 SD 可用 7.22ppm 来估计。因此抽得数之和的 SE 是$\sqrt{5}\times 7.22=16.14$ppm，并且平均数的 SE 是 16.14/5≈3.23 ppm。检验统计量是

$$\frac{77.8-70}{3.23}=2.4$$

换句话说，样本平均数高出原假设下的期望值大约 2.4 个 SE。现在，在正态曲线下，2.4 的右方面积少于 1%，因而看来有较强的证据否定原假设。

但是，这一程序却忽视某些东西。测量值的 SD 只是误差盒子的 SD 的一个估计。并且在这一例题中，测量值的个数是如此之少使得估计不大可能准确。我们必须考虑这一额外的不准确性，为此需要下列两步。

第一步。对于很多的测量值，可用测量值的 SD 来估计误差盒子的 SD。但对较少的测量值，可用较多一点的量来估计误差盒子

的SD[10]、

$$\sqrt{\frac{\text{测量值的个数}}{\text{测量值的个数}-1}}\times \text{测量值的 SD}$$

这个量缩写为SD^+

在例题中，测量值的个数为5，SD为7.22ppm。因此，$SD^+\approx\sqrt{5/4}\times 7.22=8.07$ppm。于是，用平常的方法计算SE。和的SE是$\sqrt{5}\times 8.07\approx 18.05$ppm；平均数的SE是18.05/5=3.61ppm。检验统计量成为

$$\frac{77.8-70}{3.61}\approx 2.2$$

第二步。下一步是求P值。对很多的测量值，用正态曲线就能找出比值。但对少数几个测量值，必须采用一条不同的曲线，称作学生氏曲线。如所见到，学生氏曲线上的P值。大约为5%：比正态曲线上的1%多了许多。

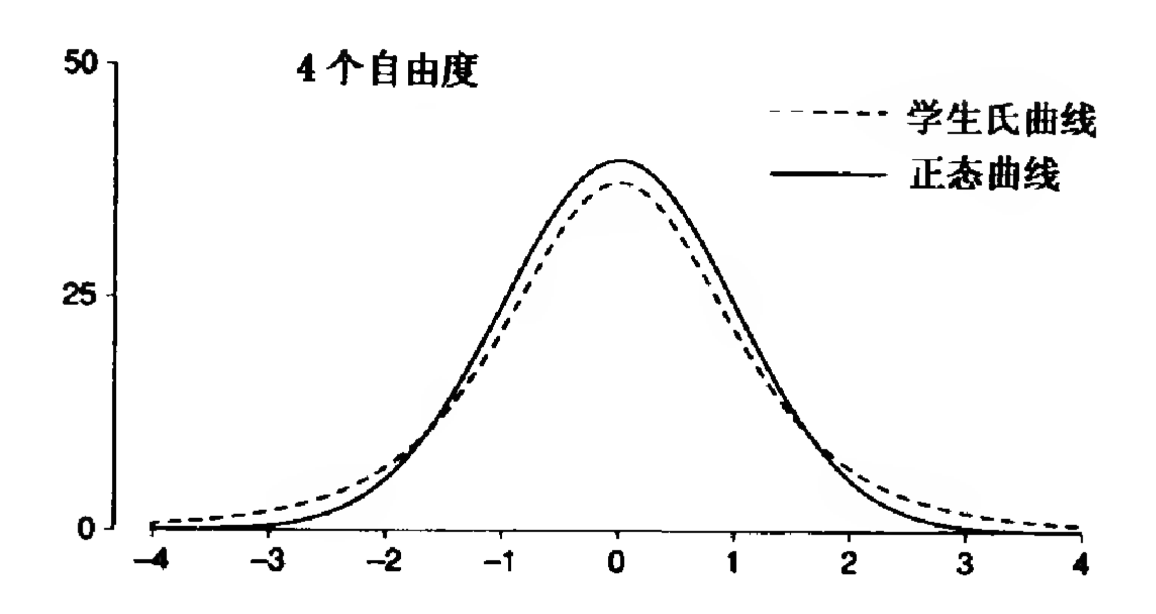
应用学生氏曲线需要做一些工作。实际上，对每一个自由度数，有一条学生氏曲线。在本书中，

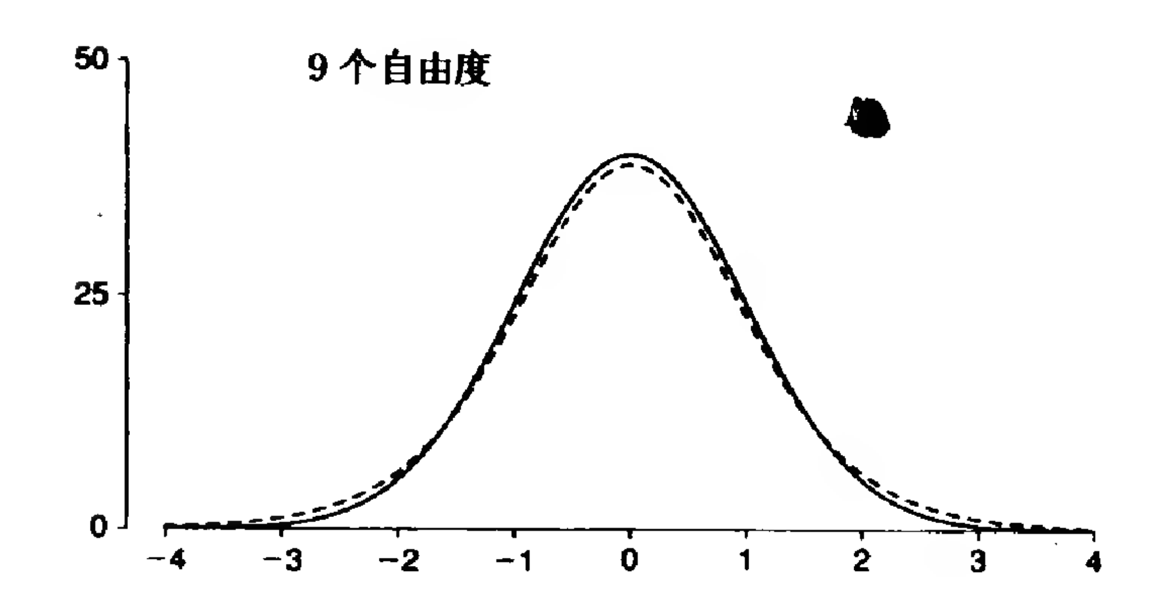
自由度＝测定数的个数－1。

（"自由度"这个词有些怪诞，但却可根据下述理由想象。平均数的SE依赖于测量值的SD而SD又反过来依赖于距离平均数的偏差。而偏差之和必须为零，因而它们不能都自由变动。于是和为零的限制消除一个自由度。）

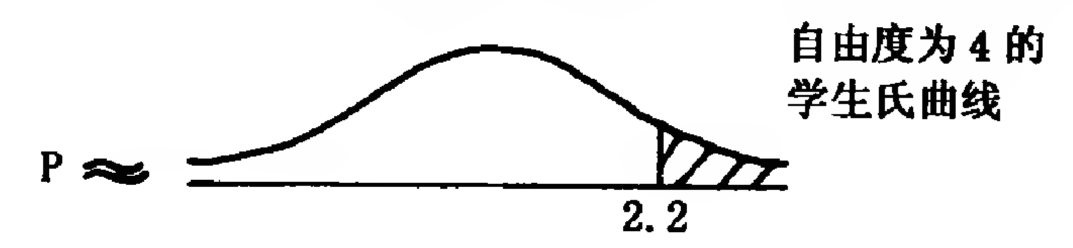
图1给出自由度为4和9学生氏曲线，用正态曲线并列在一起为比较之用。学生氏曲线看上去很象正态曲线，就是中间没有正态那么堆得高，两侧铺出得稍厚一些，当自由度愈来愈增大，学生氏曲线愈来愈接近于正态，这就反映这样的事实，测量值的SD愈来愈接近于误差盒子的SD。这些曲线都以0为对称中心，并且每条曲线下的总面积都是100%[11]。

图 1 学生氏曲线。虚线是 4 个自由度(上部)或 9 个自由度(下部)的学生氏曲线。实线是正态曲线,作比较之用。





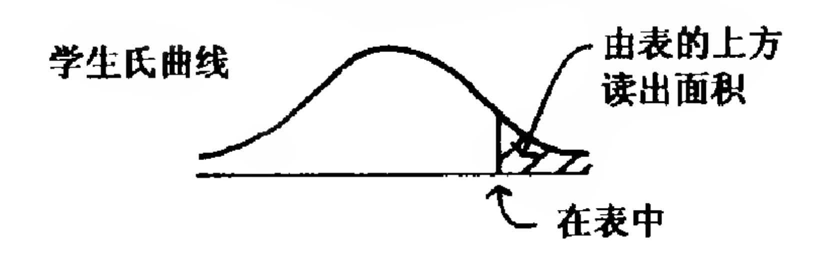
回到例题上来,5 个测量值的自由度是 5－1＝4。为了求 P 值,我们需要知道自由度为 4 的学生氏曲线下,2.2 右方的面积。



借助本书附录中一个特殊的表。我们就能找到 P 值。我们把这个表的一部份,重印在此,记为表 1。我们以自由度来标志各行。让我们沿着自由度为 4 的那行由左向右看。在顶端标志着 10% 的

那一列，我们发现 1.53。这就意味着在自由度为 4 的学生氏曲线

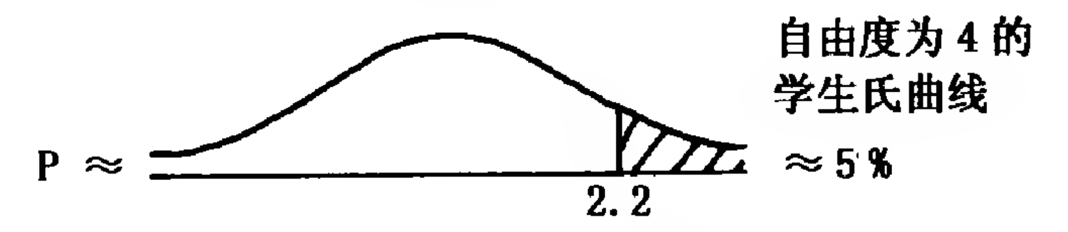
表 1　简略的 t 表



自由度	10%	5%	1%
1	3.08	6.31	31.82
2	1.89	2.92	6.96
3	1.64	2.35	4.54
4	1.53	2.13	3.75
5	1.48	2.02	3.36

下，1.53 右方的面积等于 10%。表上其他各个数字可以类似地读出。

在例题中，自由度为 4。及 t 值为 2.2。从表 1 见到在学生氏曲线下，2.13 右方的面积等于 5%。因此，2.2 右方的面积必然是略小于 5%　P 值大约为 5%。



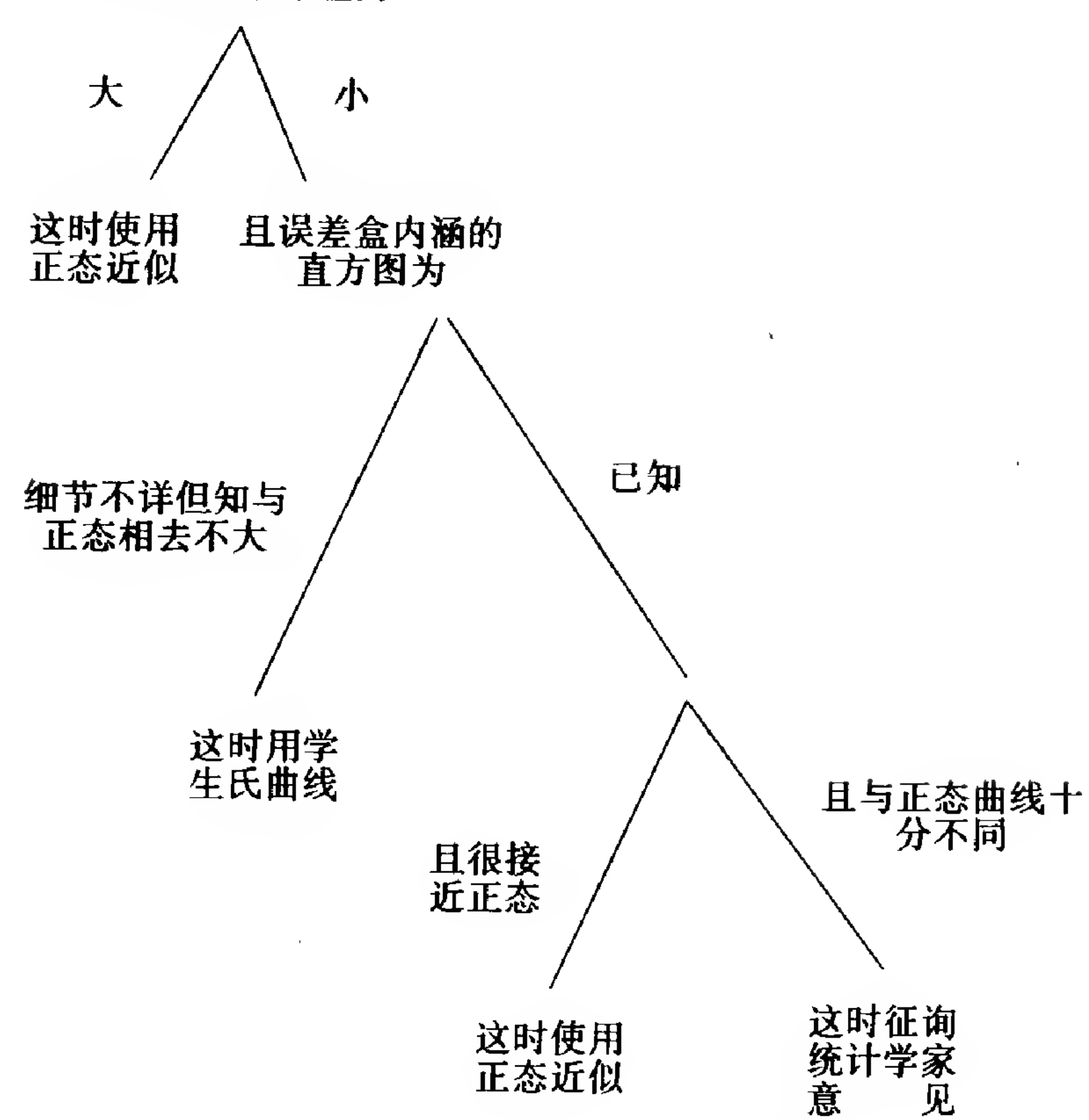
虽然不十分强烈，这一证据是违反原假设的。本题完毕。

学生氏曲线应该用在当

- 数据象从一个盒子中所抽得的数。
- 盒子的 SD 为未知。
- 观察个数少，因而盒子的 SD 不能准确地由数据估计出来。
- 盒子内涵的直方图看来与正态曲线并不相差太远。

对大量的测量值（比如说，25 个或更多一点）通常好用正态曲线。如果盒子的 SD 为已知，且盒中内涵服从正态曲线，那就可用

正态曲线，即使是小样本[12]。



例 1. 在另一天，关于量标气体的 6 个读数为

72　79　65　84　67　77

问这台机器严格地符合标准吗？还是测量值显现偏差？

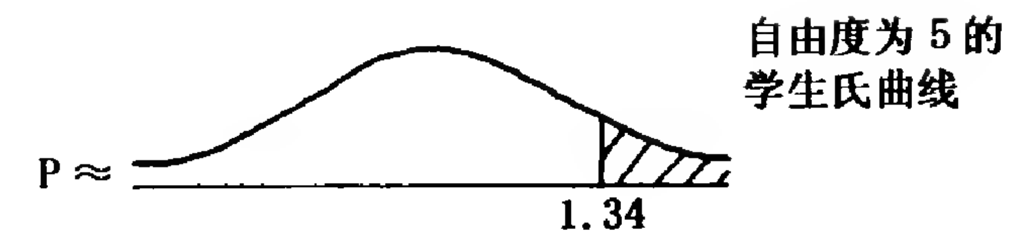
解：模型与以前的相同。新测量值的平均数为 74ppm，并且它们的 SD 为 6.68ppm，由于观察仅有 6 个，误差盒的 SD 应该用数据的 SD^+ 来估计而不是用 SD 来估计，SD^+ 等于 $\sqrt{6/5}\times 6.68\approx$ 7.32ppm，因此平均数的 SE 是 2.99，这时

$$t=\frac{74-70}{2.99}\approx 1.34$$

为了计算 P 值，我们用

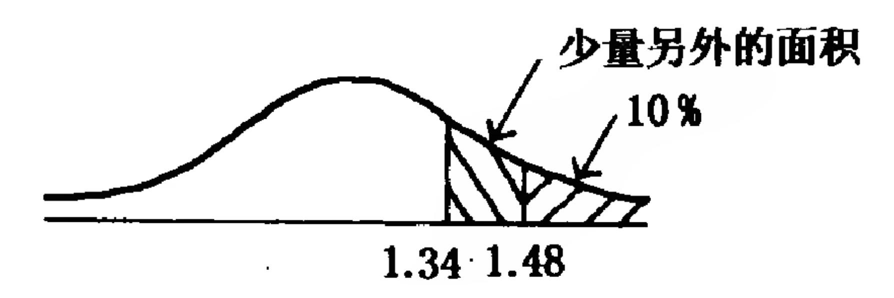
$$自由度=6-1=5$$

的学生氏曲线而不用正态



由表 1，我们查到自由度为 5 学生氏曲线下，1.34 右方的面积比 10%略为大一点，不见得有多大呈现偏性的证据。这台机器随时好用。

有关 10%的推理：由表，我们见到 1.48 之右的面积确好是 10%，而 1.34 正好在 1.48 的左边，所以 1.34 之右的面积比 10%稍微大一点。



习题 F

1. 求自由度为 5 的学生氏曲线下
   (a) 2.02 右方的面积
   (b) −2.02 左方的面积
   (c) 介乎−2.02 和 2.02 之间的面积
   (d) 2.02 左方的面积
2. 自由度为 2 的学生氏曲线下，4.02 右方的面积小于 1%，介乎 1%和 5%之间，大于 5%

   挑选一个答案，并予解释
3. 是或非，并予解释。为对 4 个测量值做一次 t 检验，使用自由度为 4 的学生氏曲线
4. 下列每一组（假设的）数据集表示关于量标气体的一些读数。假定为高斯模型，其误差服从正态曲线。然而偏性可能出现。对每一种情况作一 t 检验以察看机器是否严格地符合标准。在一种情况中 t 检验是不可能的，哪一种

情况，并说明为什么？

(a) 71,68,79

(b) 71,68,79,84,85,69

(c) 71

(d) 71,84

5. 在对一台新的分光光度器进行标定时，我们不清楚误差是否服从正态曲线，或者甚至高斯模型是否适用，在两种情况中，这些假设必须拒绝，哪两种？并说明为什么？这些数字是关于量标气体的重复测量值。

(a)71,70,72,69,71,68,93,75,68,61,74,67

(b)71,73,69,74,65,67,71,69,70,75,71,68

(c)71,69,71,69,71,69,71,69,71,69,71,69

6. 对一检验砝码进行一长系列的测量得出平均数为 10 克之上 253 微克，并且它们的 SD 为 7 微克。高斯模型认为是可用的并且偏差可忽略不记，在这一点上，天平必须重新组装，但这将引入偏差并改变误差盒的 SD。应用重装后的天平，对检验砝码 10 次测量，给出平均数为 10 克之上 245 微克，且 SD 为 9 微克。问有无引入偏差？还是这就是机会变异？(你可以假定误差服从正态曲线)

7. 对一检验砝码测量了几千次得出平均数为 1 千克之上 512 微克，且它们的 SD 为 50 微克。然后清洗砝码，下一个 100 次测量平均数为 1 千克之上 508 微克，且它们的 SD 为 52 微克，很明显，砝码轻了 4 微克。难道这不是机会变异吗？(你可以假定高斯模型但无偏差)

(a)把原假设和备择假设写成有关盒子模型的陈述。

(b)你会估计盒子的 SD 为 50 或 52 微克吗？

(c)计算 Z 和 P。

(d)砝码变轻了吗？如果是的，轻多少？

这些习题的答案在第 724—725 页上。

7. 复习题

复习题可能包括以前各章的内容

1. 是或非，并解释

(a)一个检验的 P 值等于它的观察显著水平

(b)备择假设是另一种解释结果的方式，它说明差异是由于机会而引起。

2. 一个完全平衡的轮盘(赌)在长期赌博中应在38次中出现红色数字18次。为了检验它的轮盘，一个赌场记录了3 800次赌局，发现1 890次红色数字。问红字是否太多？或者是机会变异吗？

(a)把原假设和备择假设写成有关一个盒子模型的陈述。

(b)计算Z和P

(c)红字是否太多？

3. 一种植物只开蓝花和白花。按照一种遗传模型，某种杂交植物的后代相互独立地有75%的机会开蓝花，和25%机会开白花。从这样一种杂交品种收获了200颗种子，其中142株开蓝花。问这与模型相一致否，回答是或否，并作简短解释。

4. 一个900学生的大班分成30人一组，由助教率领上课。大考后全班平均数为62分，SD为20分。然而有一组平均数为57分。助教争辩说：

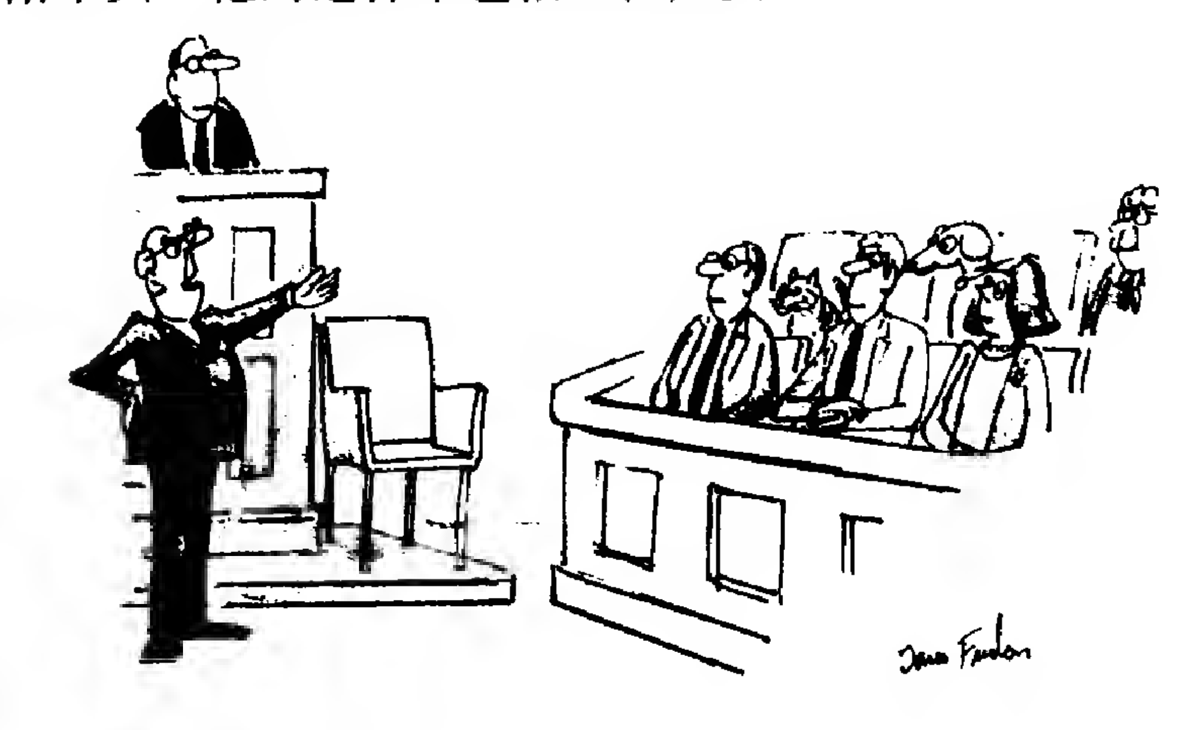
如果你随机地从大班中挑选30个学生，有很好的机会他们大考的平均分数低57分。这就是我所遇到的情况——机会变异。

问这是一个好的辩护吗？回答是或否并作简短解释。

5. 全国的数据表明(美国)大学一年级学生平均每周花7.5小时参加社交聚会[13]。一位行政人员不相信这些数字适用于她所在的有3 000个一年级学生的大学。她选择100个一年级学生的简单随机样本，并询问他们。平均来说，这些学生每周花6.6小时参加聚会，其SD为9小时。问6.6与7.5的差异是实际上存在的吗？着手回答问题之前，先用盒子模型提出原假设和备择假设。

6. 1969年，Apock博士出庭Boston联邦法院听候Ford法官审讯，被指控的罪名是阴谋破坏军役法。“在全体被告中，Apock

博士曾对数百万的母亲提供明智而又受欢迎有关生孩子的忠告，所以他希望妇女参加陪审团”[14]。全体陪审员是由一张法定名单中抽选出来的，或者由书记员从一组 350 人的小组中挑选出来的。这次的法定名单仅包含 102 女人，虽然这一辖区内可选为陪审员者过半数为妇女。在下一阶段，正式挑选陪审员来审讯此案时，Ford 法官从 350 人中挑选 100 位可能的陪审员。他的选择中包括 9 位妇女。



“尊敬的法官先生，本检察院反对这一陪审团。”

(a)350 人是从一很大的含有超过 50%妇女的母体中随机地选得的，问有多大可能取得的一个样本中含有 102 位或更少的妇女？

(b)从一群包含 102 女人和 248 男人中，随机地但不放回地抽取 100 人。问有多大可能抽得的样本中包含 9 位或更少的女人？

(c)你得出什么结论？

7. I. S. Wright 和他的同事做了一次关于抗凝剂治疗冠心病效果的临床试验[15]。对那些一月中的单日被接纳入院的合格病人给予治疗；而在双日接纳的合格病人则作为对照。以总数计，有 580 名病人在治疗组和 442 名在对照组。一位观察者

说，

因为关于去处理组或者对照组的单—双日式指派是客观且公正的。因此它恰如扔一个硬币那样地有效。

你同意还是不同意？作简短的解释。假定试验是在含有 30 天的月份里进行。

8. 书店喜欢教育，因为全国的数据表明 75%的学院毕业生在过去的六个月内读完一本书，相比较，年龄 18 岁和以上的人中只有 50%在过去的六个月内读完一本书。数据还表明对于 18 岁和 18 岁以上的人来说，全国性平均教育水平是完成 13 年正规学校教育，SD 为 4 年[16]。

   一家书店在某县做一市场调查，取 1 000 名 18 岁及 18 岁以上人的简单随机样本。他们发现平均教育水平为 14 年，且 SD 是 5 年。在样本和全国之间在教育水平上的差异能用机会变异来解释吗？如果不能，你能给出什么样的其它解释？

9. 1965 年 11 月 9 日，纽约城灯火熄灭并持续了一天—— 被称为"大停电"(Great Blackout)。九个月以后，报纸提出纽约经历了一次婴儿激增。下面的表给出了集中在大停电以后 9 个月又 10 天的前后 25 天期间内每一天所生的婴儿数[17]。这些数平均为

1966 年 8 月 1—25 日纽约的出生数

日期	星期	出生数	日期	星期	出生数
1	Mon.	451	15	Mon.	451
2	Tues.	468	16	Tues.	497
3	Wed.	429	17	Wed.	
4	Thur.	448	18	Thur.	
5	Fri.	466	19	Fri	458
6	Sat.	377	20	Sat.	429
7	Sun.	344	21	Sun.	434
8	Mon.	448	22	Mon.	410
9	Tues.	438	23	Tues.	351
10	Wed.	455	24	Wed.	467
11	Thur.	468	25	Thur.	508
12	Fri.	462			426
13	Sat.	405			
14	Sun.	377			

436。对纽约而言这不算出乎寻常地高。可是数据有一个有趣的花样。3 个星期天平均仅为 357。从此表中随机地选出 3 天其平均数为 357 或更少的可能性怎样？你作出什么样的推断？

10. 在美国，有两个关于犯罪率的全国统计来源：(i)联邦调查局的 Uniform Crime Reporting Program，它就实质上覆盖了 100% 居民的管辖范围内警方机构所得到报告的所有犯罪公告了概要；(ii)The National Crime Survey（全国性犯罪调查），建立在与一个全国性的家庭概率样本的访问会见基础上[18]。

    1986 年，样本中 6.3% 的家庭告诉访问员他们至少经历过一次盗窃。那一年，联邦调查局报告了每 1 000 户家庭中 13 户，或 1.3% 的盗窃率。这个差异能用机会误差解释掉吗？如果不能，你怎样解释它？你可以假设调查基于 50 000 户家庭的一组简单随机样本。

11. 自十三世纪初以来，由大不列颠皇家造币厂铸造的钱币以一组样本为基在一个检验钱币样品箱中进行检验[19]。假使在 1799 年，检验程序如下：从该年大量生产的畿尼（英国旧金币）中随机地选出 1 000 枚放入样品箱（一个正式的匣子）内，然后进行秤量。造币厂厂长被容许有一误差幅度——称之为（制造硬币的）公差——是根据当时的制造容限来确定的。

    1799 年，一畿尼应重 128 谷（一磅有 5 760 谷）：因此在畿尼样品箱中的 1 000 枚畿尼应重 1 000×128 谷。公差是该量的 1/200 倍，或 640 谷。如果样品箱中钱币的实际重量与标准相差多于公差，造币厂厂长则面临严厉的处罚。

    (a) 假如在 1799 年，造币厂厂长制造畿尼重 128 谷，允许有 200 分之一的随机误差。他从钱币样品箱的检验中逃生的机会有多少？

    (b) 如果他调整机器使得制造的畿尼平均来说重 127.5 谷，他逃出的机会又有多少？（随机误差依然相同。）

    (c) 如果造币厂厂长用调整为 127.5 谷的机器制造了 100 000

枚鑯尼,他将侵吞____土____左右谷的金,(只要他从钱币样品箱检验中逃生,他将获取这些不义之财。)

12. 心理环境影响脑子的分析吗? 这个问题由 Mark Rosenzweig 和其他人通过实验得到研究[20]。研究的对象来自一些遗传学纯种的老鼠中,从每一窝老鼠中,随机地选择一只去处理组,另一只去对照组。两个组得到完全一样品质的食物和饮料——尽它们所需。但是处理组中的每一只老鼠与其他 11 只同住一大笼子,提供每天更换的玩具。对照组中的老鼠孤立地生活,没有玩具。一个月以后,杀掉实验动物并进行解剖。

实验动物的皮质重量(毫克)。处理组(T)有丰富的环境,对照组(C)有被剥夺的环境。

实验 T	#1 C	实验 T	#2 C	实验 T	#1 C	实验 T	#2 C	实验 T	#1 C
689	657	707	669	690	668	700	662	640	641
656	623	740	650	701	667	718	705	655	589
668	652	745	651	685	647	679	656	624	603
660	654	652	627	751	693	742	652	682	642
679	658	649	656	647	635	728	578	687	612
663	646	676	642	647	644	677	678	653	603
664	600	699	698	720	665	696	670	653	593
647	640	696	648	718	689	711	647	660	672
694	605	712	676	718	642	670	632	668	612
633	635	708	657	696	673	651	661	679	678
653	642	749	692	658	675	711	670	638	593
		690	621	680	641	710	694	649	602

平均来说,对照组动物较重且有较重的脑髓,也许因为它们吃得较多以及活动得较少。但是,处理组一致地有较重的皮质("灰白质",或大脑的思考部分)。该实验被重复许多次;头次试验的结果给出如下:T 意指处理,C 表示对照。每一排涉及一对动物。在第一对中,处理组的动物有重 689 毫克的皮质;对照组的具有较轻的皮质,仅重 657 毫克。等等。

分析数据的两种方法将以习题的形式提出。两种方法均成对考虑,这是数据的关键特性。成对的结果来自于事实:两个动

物来自于同一窝。

(a)总共有多少对？这些对中有多少对的处理组动物有较重的皮质？

(b)假定处理没有效应，因此对中的每一只动物关于对之间为独立地有 50 对 50 的机会具有较重的皮质。在这假设下，一位研究人员得到如同 Rosenzweig 一样多或更多的对，其处理组动物有较重皮质的可能性是多少？你作出什么样的推断？

(c)对于每一对动物，计算“处理—对照”皮质重量的差。求出所有这些差的平均数和 SD。原假设讲这些差象从一个平均数是 0 的匣子中作随机有放回的抽取：处理没有效应。对该假设作了检验。你作出什么样的推断？

(d)为保证分析的有效性，作如下预防措施。“每一组同窝老鼠的大脑解剖和分析是紧挨着接二连三地进行的，但是随机的且只能根据代码来认同，这样使得做解剖的人不知道老鼠来自哪一笼。”简短地详注下述问题：这种举措的目的意义是什么？它是个好主意吗？

8. 小结

1. 显著性检验查明观察到的差异是实际存在的(备择假设)还是只是机会变异(原假设)这个问题。

2. 合理的显著性检验需要关于数据的一个盒子模型。原假设必须转写为关于盒子模型的陈述。通常，备择假设也需要这样做。

3. 检验统计量用来度量在数据与基于原假设可预定的值之间的差异。Z 检验使用统计量

$$Z=\frac{\text{观察值}-\text{期望值}}{\text{SE}}$$

期望值是基于原假设计算出来的。

4. 观察到的显著性水平(也称 P，或 P－值)是得到一个检验

统计量如同观察的一样极端或者比观察的更极端的机会。该机会是在原假设为正确的这个基础上计算出来的。因此,P 并不给出原假设是正确的机会。

5. P 为小的数值是推翻原假设的证据;它们指出了在机会起作用而产生差异之外的某些东西。

6. 假设随机有放回地从一个其内容服从均值为 0 而 SD 未知的正态曲线的匣子中抽取少量票。每次抽取被加上一个未知常数以给出一测量值。原假设说这个未知常数等于某给定常数C。备择假设说未知常数大于 C。匣子的 SD 由数据的 SD^+ 来估计。于是关于抽取的平均数的 SE 可计算出来。检验统计量是

$$t=\frac{\text{抽取的平均数}-C}{SE}$$

观察到的显著性水平不是从正态曲线得到而是从某一学生氏曲线得到,它具有

$$\text{自由度}=\text{测量次数}-1$$

这种方法是 t-检验。

27
再论平均数检验

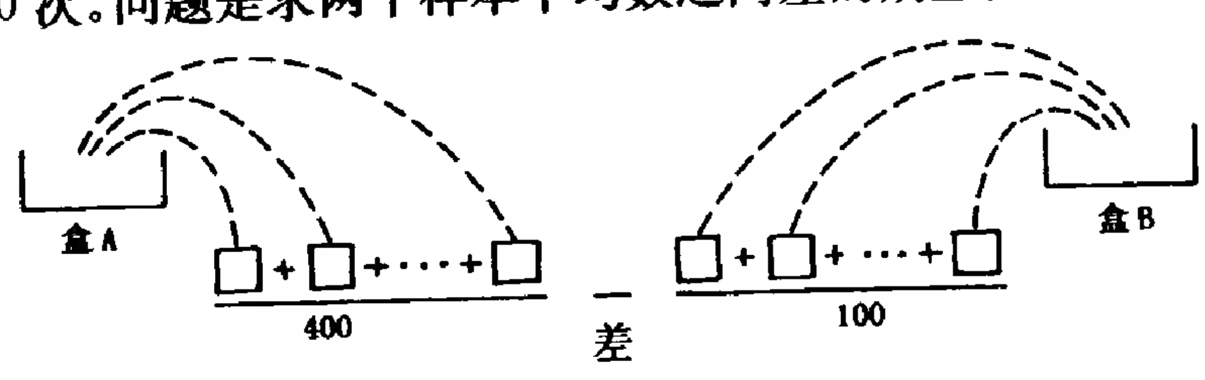
差异万岁!

1. 差的标准误差

本章打算比较两个样本的平均数。需要用到它们的差的标准误差。我们先举一个例子说明数学内容——实际例子稍后谈及。设有两只盒子 A 和 B,它们具有如下所示的平均数与 SD。

盒 A	盒 B
平均数=110 SD=60	平均数=90 SD=40

从盒 A 中随机放回地抽取 400 次,另独立地从盒 B 随机放回地抽取 100 次。问题是求两个样本平均数之间差的期望值和标准误差。



第一步分别算出各样本平均数的期望值和SE(第23章第1节):

盒A的400个抽得数的平均数=110±3左右

盒B的100个抽得数的平均数=90±4左右

差的期望值就是110−90=20,但新的问题是如何将两个SE汇总起来:

$$(110\pm3)-(90\pm4)=20\pm____?$$

一个自然的猜测是将2个SE相加:3+4=7。但是,这样做忽略了相互抵消的可能性。正确的SE可由平方根法则求得。

两个独立量间差的标准误差为$\sqrt{a^2+b^2}$,其中 • a是第一个量的SE • b是第二个量的SE

在本例中,从两个盒子中抽取是独立进行的,因此两个平均数是独立的,平方根法则适用。现在a是3,b是4,故两个平均数间的差SE为

$$\sqrt{3^2+4^2}=\sqrt{25}=5。$$

例1. 随机放回地从下面所给的盒C中抽取100次。另又独立随机放回地从盒D中抽取100次。求从盒C中抽到数字1的次数与从盒D中抽到数字4的次数间差的期望值与SE。

C [1][2]　　　　D [3][4]

解:数字1的次数约为50±5左右。数字4的次数亦约为50±5左右。差的期望是50−50=0。抽取是独立进行的,所以两个次数是独立的,平方根法则适用。差的SE为$\sqrt{5^2+5^2}\approx7$。

例2. 随机放回地从盒子[1][2][3][4][5]中抽取100次。抽得数字1的期望次数是20且SE为4。抽得数字5的期望次数也是20,SE为4。正确与否:抽得数字1的次数和抽得数字5的次数间差的SE是$\sqrt{4^2+4^2}$。

解:这是错的。这两个次数是相依的,因为当一个大时另一个必定小。平方根法则不适用。

习题 A

1. 盒 A 的平均数是 100,SD 是 10。盒 B 的平均数是 50,SD 是 18。现从盒 A 随机放回地抽取 25 次,又独立地从盒 B 随机放回地抽取 36 次。求从盒 A 抽得数的平均数与从盒 B 抽得数的平均数间差的期望值和标准误差。
2. 一枚硬币抛 500 次。求前 400 次中头像百分数与后 100 次中头像百分数间差的期望值和 SE。
3. 某盒中装有 5 000 张标有数字的票子,平均数是 50,SD 是 30。随机不放回地抽取 200 张。正确与否,并解释:头 100 次抽得数的平均数与第二批 100 次抽得数的平均数间差的 SE 近似为$\sqrt{3^2+3^2}$。(提示:若抽取是有放回时将如何?)
4. 假设盒子中只装 200 张票,重复习题 3。抽取是不放回的。
5. 从盒 A 中随机放回地抽 100 次,这些抽得数的平均数是 51,它们的 SD 是 3。独立地从盒 B 中随机放回地抽 400 次;抽得数的平均数是 48,它们的 SD 是 8。某人声称这两只盒子有相同的平均数。你是怎么想的?

这些习题的答案在第 725 页上。

2. 比较两个样本的平均数

学校的成绩变得较差了吗?每年全国教育发展评估(NAEP)都要对一个由 17 岁学生组成的全国性样本实施若干科目的考试[①]。数学考试在 1973 年进行了一次,1982 年又进行一次;平均得分由 55.0 降到 51.8。差是 3.2 分。这是实际存在的,还是仅为机会变异?

可以用 Z—检验,但是 SE 的计算比第 26 章更复杂。在那里,一个样本平均数与某客观标准相比较。而这里,有两个样本,它们的平均数的差是要讨论的问题:

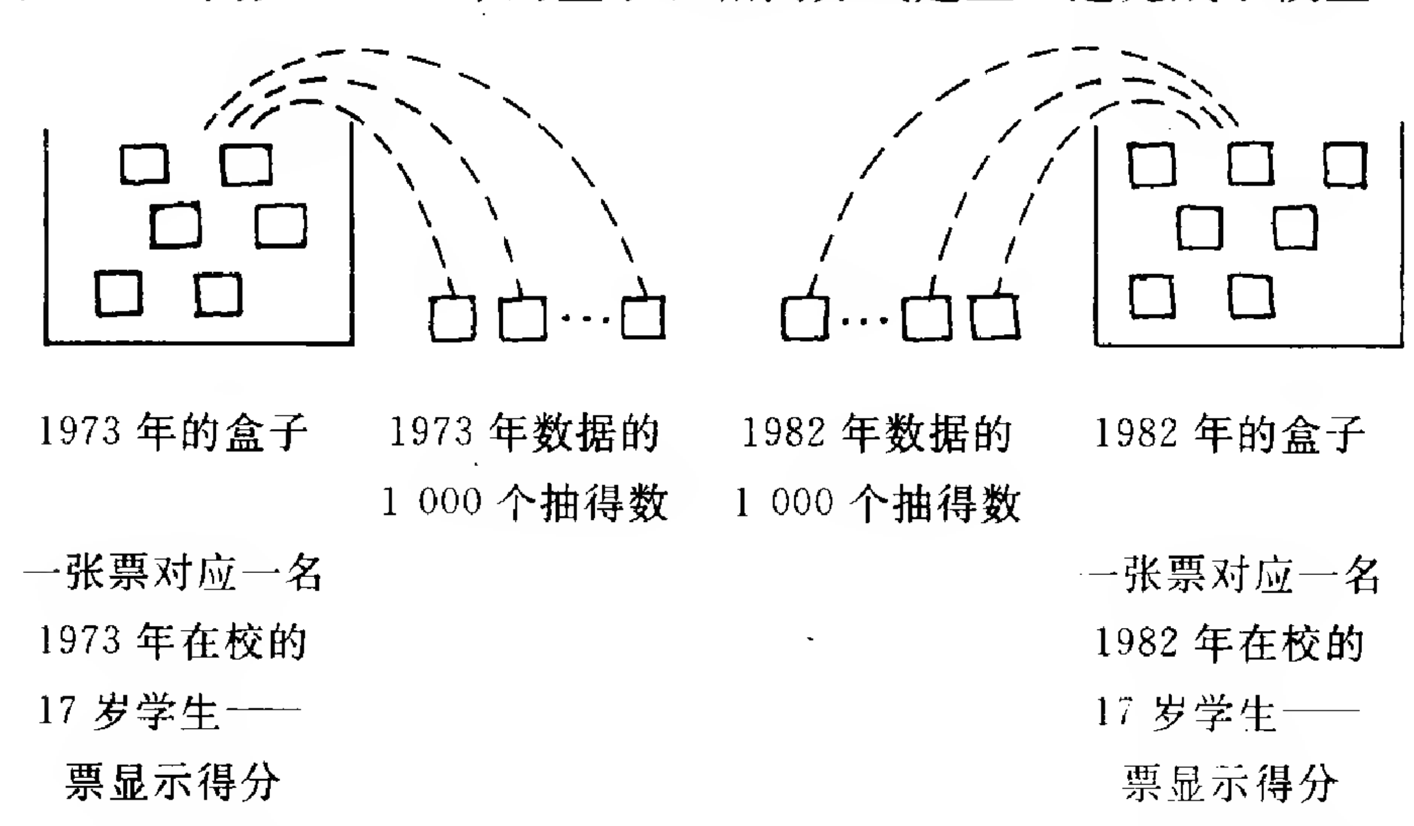
1982 年样本的平均得分—1973 年样本的平均得分

两个平均数都受到机会变异,差的 SE 必须考虑到这一点。第一节的方法可以使用。

为计算标准误差,需要一个盒子模型,另外样本的设计也有关

系的。这相当复杂;这里给出一个简化形式。假设,在1973年和1982年,考试是对一个当时在校注册的1 000名17岁学生的全国性简单随机样本进行。

在这种设计下,模型是比较简单的。必须用两只盒子,两个考试年各对应一个。1973年的盒子里有成百万张票——每张对应一个在校注册的17岁学生。票上记载着该学生所得分数,如果他或她参加了NAEP数学考试。1973年的数据就象从盒子中随机抽出的1 000个数。1982年的盒子以相同方式建立。这完成了模型。



原假设称两只盒子的平均数相等。在这基础上,样本平均数间差的期望是0,观察到的差只是反映了抽取的侥幸;美国的年青人没有走下坡路。备择假设称1982年盒子的平均数小于1973年盒子的平均数——数学得分确实下降,并且这就是为什么两个样本平均数不同的原因。两样本Z——统计量将帮助在这两个假设间做出选择:

$$Z=\frac{\text{观察差}-\text{期望差}}{\text{差的SE}}$$

我们先看分子。观察差是51.8−55.0=−3.2。在零假设下,两个样本平均数间的差期望为0,故分子是

$$-3.2-0=-3.2$$

现在来看分母。一次看一个样本。1973 年，1 000 个考试成绩的 SD 为 17.2。因此 1973 年盒子的 SD 估计为 17.2；1973 年 1 000 个考试成绩的和的 SE 为$\sqrt{1\ 000}\times 17.2\approx 544$；平均数的 SE 为$\frac{544}{1\ 000}\approx 0.54$。1982 年，SD 为 16.9，1982 年平均数的 SE 为 0.53。差的 SE 可以用上节的方法计算，因为样本是独立的：

$$\sqrt{0.54^2+0.53^2}\approx 0.76$$

最后，

$$Z\approx -3.2/0.76\approx -4.2$$

换句话说，在原假设下，1982 年与 1973 年间的差约低于期望值 4.2 个 SE——所取机会变差达到远离中心的范围。我们拒绝原假设，保留备择假设，即差是实际存在的。

两样本 Z——统计量由 • 两个样本的容量 • 两个样本的平均数 • 两个样本的 SD 计算出来。检验假定这是两个独立的简单随机样本。

在 NAEP 例子中，样本都足够地大以致于每个样本的平均数的概率直方图遵循正态曲线。于是 Z 也遵循正态曲线。

两样本 Z——检验也可用于百分数，如下例所示。

例 3. 某规模较大的大学取了一个男生简单随机样本，和另一个女生简单随机样本。样本容量分别是 200 和 300。结果是，107 个样本男生经常使用个人计算机，对比于 132 个样本女生：53.5%对 44.0%。百分数间的差是实际存在的，还是机会变异？②

解：存在一只盒子，装着对应该大学全体男生的票子；如果该生经常使用 PC 则相应票子标上 1，否则标 0；男生的数据就象取自这只盒子的 200 个抽得数。对女生可以如法炮制，抽取 300 次。以模型的说法，原假设称这两只盒子中 1 的百分数相同。备择假设

称这两只盒子中的百分数不同。

为做Z—检验，我们必须对两个样本百分数间的差赋于一个SE。一次取一个样本。使用PC的男生人数的SE估计为

$$\sqrt{200}\times\sqrt{0.535\times0.465}\approx7$$

百分数的SE为

$$\frac{7}{200}\times100\%=3.5\%$$

类似地，使用PC的女生百分数的SE为2.9%。差的SE为

$$\sqrt{3.5^2+2.9^2}\approx4.5(\%)$$

在原假设下，期望差为0.0%。观察到的差为53.5−44.0=9.5(%)。故检验统计量为

$$Z=\frac{\text{观察差}-\text{期望差}}{\text{差的SE}}\approx\frac{9.5\%-0.0\%}{4.5\%}\approx2.1$$

再一次，我们拒绝原假设[3]。

习题B

1. “两个样本的平均数间的差只是由于机会的缘故吗?”为有助于回答这个问题，统计学家们使用______ ______Z—检验。填空并简短解释。
2. 健康检查调查的第三轮使用一个12到17岁青少年的全国性概率样本。调查的目的之一是估计青少年文盲的百分数[4]。设计了一次考试以测量阅读和写作能力。它包含七段短文，每段附三个问题，象下面的那样：

   门口有脚步声和敲门声。里面的每一个人迅速站起身。仅有的响声是炉灶上水壶烧开所发出的声音。又是一下敲门声。没有人移动。能听到门外的脚步声离去。

   · 房内的人

   (a)躲在炉灶后面

   (b)迅速站起身

   (c)跑向门口

   (d)大声笑

   (e)开始哭叫

· 房内唯一的响声是什么？

(a)人们交谈

(b)鸟儿歌唱

(c)一只水壶在烧开

(d)一只狗在吠

(e)一个男人在喊

· 敲门的人最后

(a)步入房内

(b)在门外坐下

(c)大声呼救

(d)离去

(e)砸坏门

这次考试被设计在四年级阅读水平上，若能正确回答半数以上的问题，则确定为有阅读和写作能力。

男性与女性在该考试中的成绩间出现某种差异：7%的男性有阅读和写作能力，对比 3%的女性。这差异是实际存在的，还是机会变异的结果？

你可以假设调研员取了一个 1 600 名男性青少年的简单随机样本，和另一个独立的 1 600 名女性青少年的简单随机样本。在回答问题前用盒子模型的说法提出原假设和备择假设。

3. 健康检查调查的第二轮使用一个 6 到 11 岁儿童的全国性概率样本。调查的目的之一是研究智力测验中儿童的得分与他们的家庭背景之间的关系[5]。所采用的一项智力测验是 WISC 词汇量级别。它由儿童必须解释的 40 个单词组成。回答对的给 2 分，部分回答对的给 1 分。得分出现明显地与父母的收入和受教育水平相关。另外，测验的得分与父母所住社区的类型之间也存在相关。譬如，大城市的儿童平均测验得分 26 分，其 SD 是 10 分；但农村的儿童平均仅 25 分，且 SD 也是 10 分。这个差可以解释为机会变异吗？你可以假设调研员取了一个 400 名大城市儿童的简单随机样本，和另一个独立的 400 名农村儿童的简单随机样本。在回答问题前用盒子模型的说法提出原假设和备择假设。

4. 将两个样本的容量 400 都改成 1 000 后重复习题 3。

5. 第 26 章的复习题 12 叙述了一项试验，其中 59 只动物被放进处理组（丰富了的环境），另外 59 只供做对照。处理组的平均大脑灰质重是 683 毫克，

SD是31毫克。对照组的平均大脑灰质重是647毫克,SD是29毫克。某人建议做两样本Z—检验:

处理组重量之和的SE$\approx\sqrt{59}\times31\approx238$毫克

处理组重量的平均数的SE$\approx238/59\approx4.0$毫克

对照组重量之和的SE$\approx\sqrt{59}\times29\approx223$毫克

对照组重量的平均数的SE$\approx223/59\approx3.8$毫克

差的SE$\approx\sqrt{4.0^2+3.8^2}\approx5.5$毫克

$Z=(683-647)/5.5\approx6.5$



统计理论说明什么?

6. 全国家庭滥用药物调查于1974年进行了一次,而后又于1985年再进行一次[6]。

(a)18岁到25岁的人中,吸食大麻的百分数由25.3%降为21.9%。这个差是实际存在的,还是机会变异?

(b)18岁到25岁的人中,吸食可卡因的百分数由3.1%上升为7.7%。这个差是实际存在的,还是机会变异?

你可以假设这些结果是基于两个容量各为700的独立的简单随机样本。

7. 公立大学的新生为了获取津贴平均每周工作12小时,SD是10.5小时;在私立大学,平均数是9.2小时,SD是9.9小时。假设这些数据是基于两个容量各为1 000人的独立简单随机样本[7]。平均数之间的差归因于机会吗?若不是,别的什么可以解释它?

8. 某大学取了一个132名男生和279名女生的简单随机样本,53%的男生和48%的女生汇报在调查周中工作超过8小时。为了查明在百分数中的差是否统计显著,调研员从计算$Z=(53-48)/0.53$着手。出了什么差错吗?

这些习题的答案在第726页上。

3. 实验

第1节的方法也可用来分析某类实验数据,在这些实验中,调研员随机选取一些实验对象接受处理“A”,而另一些接受“B”。譬

如，在脊髓灰质炎疫苗现场实验中，处理A是疫苗；处理B，是给对照组的安慰剂(第1章)。我们先用一个例子讲解方法的技巧，然后再解释为什么方法可行。

例4。在一次关于维生素C的小型诊所实验中，有200名实验对象。随机指派半数对象到处理组(日服维生素C 2 000毫克)，另半数到对照组(服2 000毫克安慰剂)。在整个实验期间，处理组平均患感冒2.3次，SD是3.1次。对照组稍差一些：他们平均患感冒2.6次，SD是2.9次。平均数中的差异统计显著吗？

解：两个平均数之间的差是−0.3，你需对这个数赋予一个标准误差。假装认为你有两个随机放回抽取的独立样本。处理组之和的SE是$\sqrt{100}\times 3.1=31$；处理组平均数的SE是31/100=0.3。类似地，对照组平均数的SE是0.29。差的SE是

$$\sqrt{0.31^2+0.29^2}\approx 0.42$$

原假设称维生素C无作用。在这基础上，差的期望值是0.0。观察差是−0.3。故

$$Z=\frac{\text{观察差}-\text{期望差}}{\text{差的 SE}}=\frac{-0.3-0.0}{0.42}\approx -0.7$$

易见差应归因于机会：稍许多一点容易患感冒的人被指派到对照组[8]。

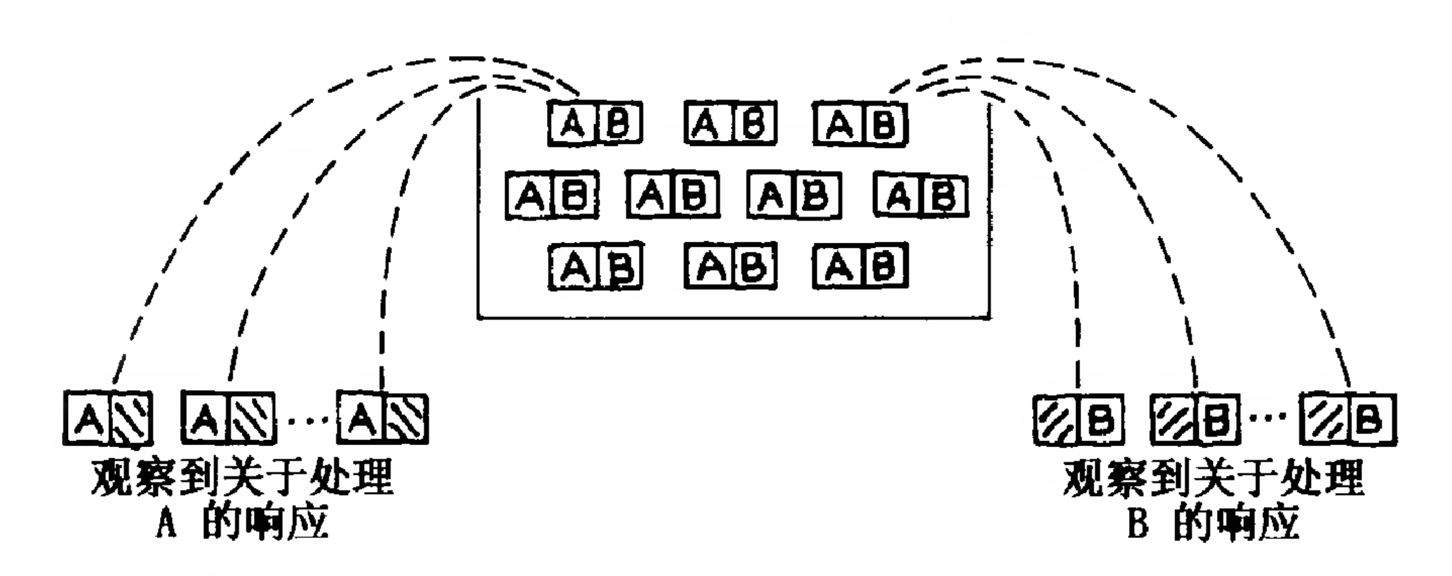
现在，看看幕后。在解例题中，要求假装认为处理组样本和对照组样本是独立地、随机放回地从两只盒子中抽取。但是，实验并不按此方式进行。有200名实验对象；100名被随机——不放回——选出服用维生素C；另100名服用安慰剂。因此，抽取是不放回的，样本是相依的。(譬如，某一实验对象可能很容易患感冒；若这个对象在维生素C组，他就不可能在安慰剂组；因此，这种指派稍许影响了两个平均数。)

为什么尽管有这些问题，SE仍直接解出？答案与获取盒子模型的真相有关。别忘了，调研员们正进行一项实验。他们随机地选择一些实验对象接受处理A，另一些接受处理B。模型照例对每一

个对象对应一张票。但是，现在票上有两个数。一个表示对处理 A 的反应会是什么，另一个是对 B 的。见图 1。两个数中只有一个能被观察到，因为对一个实验对象只能施于两种处理之一。

在这模型中，随机不放回地从盒子中抽取一些票，并观察到关于处理 A 的反应。处理 A 的数据就是第一批反应。然后，随机不放回地从盒子中更多地抽取，并观察到关于处理 B 的反应。处理 B 的数据就象这第二批反应。在例 4 中，200 名实验对象的每一个不是分在维生素 C 组就是分在安慰剂组。在这种情况下，第二个样本就相当于第一个样本抽取后盒子中留下的票子。

图 1 比较处理 A 和 B 的随机化对照实验的盒子模型。每一个实验对象对应一张票子。票上有两个数：一个表示对象对处理象 A 的反应；另一个，对处理 B。两个数中只有一个能被观察到。



原假设称两种处理的平均反应相同。为了检验这一假设，调研人员通常比较两个平均数(或两个百分数)：

A 组平均反应－B 组平均反应。

这个差的 SE 是多少？例 4 的解似乎包含了早先提及的两种错误——

- 抽取是不放回的；但两个 SE 都基于有放回抽取。
- 两个平均数是相依的；但两个 SE 就象平均数是独立地那样被汇总。

若抽取次数相对于盒子中的票数是小的，则无一错误是严重

的：放回与不放回抽取之间没什么差别，平均数间的相依性也小。几乎存在两只盒子，一只是处理组的，一只是对照组的。（第 562 页上的习题 3 就是一例。）

另一方面，若抽取次数相对大时，则各种错误本身的影响可能是实质性的。譬如，当半数受试对象分派到处理组时——如例 4——校正因子将明显小于 1（第 20 章第 3 节）。因此，第一节的方法用于随机化试验的时候运气不错，稍许保守些：它倾向于稍微偏大地估计 SE。这是由于 两种错误彼此抵销：

(i) 第一种错误使 SE 增大。

(ii)第二种错误将 SE 削减下来[9]。

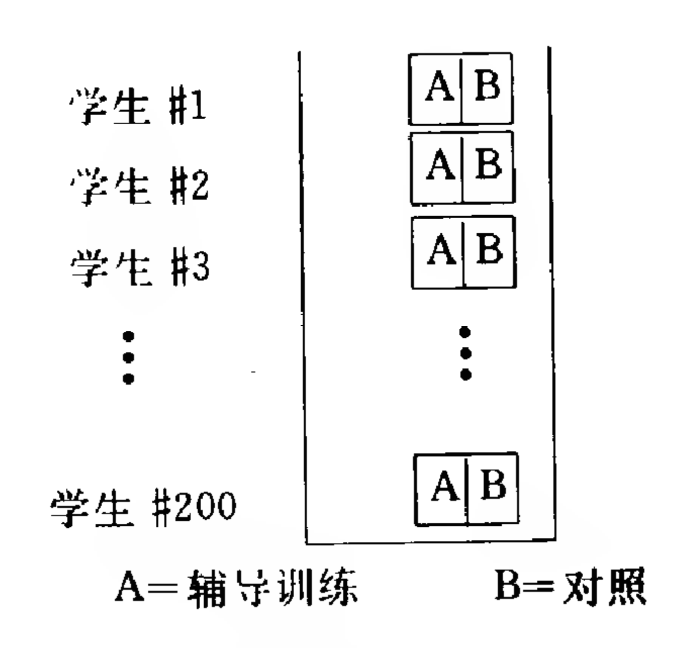
为了使得数学上可行，两个样本平均数的 SE 必须在有放回抽取基础上计算——尽管抽取是不放回的。此即所谓对两个样本间相依性的补偿。

小结：当数据来源于某随机化实验（如例 4）时，无论抽取次数是小或大，第一节的方法都可以使用。

> 有一盒票。每张票上有两个数：一个表示对处理 A 的反应。另一个则是对处理 B 的反应；两个数中仅一个可被观察。随机不放回地从盒中抽取一些票，并观察到对处理 A 的反应。然后，从留下的票子中随机不放回地抽取第二个样本，观察到对处理 B 的反应。两个样本平均数间差的 SE 可以保守地估计如下：
>
> (i)在随机放回抽取基础上计算两个平均数各自的 SE。
>
> (ii)象样本是相互独立那样将两个 SE 汇总。

习题 C

1.(假设的。)辅导训练对数学 SAT 有作用吗?一组共 200 名中学高年级生自愿作为一项实验的对象;从中随机选 100 名接受辅导训练,其余 100 名为对照。6 个月后,所有 200 名实验对象都参加数学 SAT。对这项实验建立盒子模型如下。



(a)仔细解释学生#1 的两个数表示什么。

(b)若学生#1 被指派在辅导训练组,两个数中的那一个被观察到?若这学生被指派在对照组又将如何?你可能两个数都观察到吗?

(c)辅导训练组 SAT 数学平均得 486 分;他们的 SD 是 98 分。对照组平均得 477 分;并有 103 分的 SD。辅导训练有作用吗?还是机会的作用?

2. Wheaties 是智力早餐吗?为了寻找答案,在初级统计学的某大班上进行研究;549 名学生同意参加;在期中之后,随机指派 250 人在处理组,另 249 人在对照组。处理组每周 7 天早餐都吃 Wheaties。对照组吃 Sugar Pops。

(a)期终成绩,处理组平均是 66 分,SD 是 21 分。对照组的统计数据是 59 分和 20 分。你得出什么结论?

(b)研究的哪些方面能够被"盲目地"进行?

注:Wheaties 和 Sugar Pops 是注册商标。这项研究是虚构的。

3. 本题续习题 2

(a)处理组的期中成绩平均为 61 分;SD 是 20 分。对照组的统计数据是 60 分和 19 分。你得出什么结论?

(b)如果处理组的期中平均成绩是 68 分,SD 是 21 分,而对照组的统计数

字是 59 分和 18 分。再做一遍。

4. 设例 4 中的研究，对 2 000 名实验对象重做一遍，其中 1 000 指派去维生素 C 组，另 1 000 名在对照组。又设维生素 C 组平均感冒 2.4 次，SD 是 2.9 次；而对照组的平均数是 2.5 次，SD 是 3.0 次。

(a)两个平均数间的差统计显著吗？你得出什么结论？

(b)为什么从一次研究到下一次研究平均数会发生变化？

这些习题的答案在第 727 页上。

4. 再谈实验

前面一节中所叙述的方法也可用于反应不是定量的而是定性的实验，这样票上应显示 0 和 1。本节将给出一个例子，但是在此之前，先给出一些背景材料。经济行为的标准理论设想“合理的”决定依据某些正规(或许并不现实)的规则形成。特别地，该理论指出决策者对事实作出反应，不是对提出事实的方式作出反应。另一方面，心理学家倾向于认为事实提出的方式是有关系的。经验工作偏向心理学观点[10]。

由 Amos Tversky 等人进行的一项研究涉及对外科手术或射线疗法作为肺癌备用疗法的有效性提出的信息。实验对象是 Harvard(哈佛)某暑期课程的一组 167 名医生[11]。信息以两种不同的方式提出。一些医生得到形式 A，它报导了治疗后的死亡率：

> 形式 A：每 100 个做外科手术的人，10 个将在手术中死亡，32 个将不迟于一年而死亡，66 个将不迟于 5 年而死亡。每 100 个接受射线疗法的人，无人在疗程中死亡，到一年时将有 23 人死亡，到 5 年将有 78 人死亡。

其余医生得到形式 B，它报导了存活率。

> 形式 B：每 100 个做外科手术的人，90 个将在手术中存活，68 个将存活一年或更长，34 个将存活五年或更长。每 100 个接受射线疗法的人，所有人将在疗程中存活，77 个将存活一年或更长，22 个将存活五年或更长。

两种形式包含完全相同的信息:譬如,100 名病人中 10 人在手术中死亡(形式 A),故 100 人中的 90 人将在手术中存活(形式 B)。到第五年,肺癌病人的前景是很暗淡的。

在实验中,随机从 167 名医生中选出 80 人,并给他们形式 A;余下的 87 人得到形式 B。每个医生阅后写下他或她对肺癌病人推荐的疗法。对于形式 A 的反应,80 名医生中的 40 人选择手术(表 1)。但是,对于形式 B 的反应,87 人中的 73 人选手术:40/80 是 50%,而 73/87 是 84%。提出报告的方式看来有关系。

表 1　关于数据提出方式的影响的实验结果

形式	得到该形式的人数	赞成手术人数	赞成手术百分率
A	80	40	50%
B	87	73	84%

为标准决策论辩护的某经济学家可能争辩称差只是归因于机会——原假设。基于在两种形式中呈现相同信息,一些医生将推荐手术,而另一些则选择射线疗法。但决策不能依赖于信息呈现的形式。该经济学家可能会说,仅仅由于抽取的运气,赞成手术的医生中过多的人被选中得 B 形式;他们中过少的人得 A 形式。毕竟,在 A 组中仅 80 人,而 B 组中是 87 人;因此存在机会变差的很大余地。

为评价这一争论,需要显著性检验。表 1 中两个百分率的差是 34%,你必须对这个差添上标准误差。例 4 的方法可以使用。设想你有两个随机放回地抽取的独立样本。

第一个样本是得到形式 A 的医生。由表的第一行,他们是 80 人,且 40 人赞成手术疗法。赞成手术疗法人数的 SE 是

$$\sqrt{80}\times\sqrt{0.50\times0.50}\approx4.5$$

故赞成手术疗法百分率的 SE 是 $4.5/80\times100\%=5.6\%$。第二个样本是得到形式 B 的医生。在对形式 B 的反应中赞成手术的百分

率的 SE 是 3.9%。差的 SE 是

$$\sqrt{5.6^2+3.9^2}\approx 6.8\%$$

在原假设下，两个样本中百分率间差的期望为 0.0%，观察到的差是 34%，检验统计量为

$$Z=\frac{\text{观察差}-\text{期望差}}{\text{差的 SE}}=\frac{34\%-0.0\%}{6.8\%}=5.0$$

对表 1 的结果，机会不是一个好的解释。

幕后另一瞥：这实验的盒子模型中对每个医生需要一张票。就象前面一节中那样，每张票上显示一对数[A|B]。票上的第一个数是对形式 A 反应的代号：一个医生收到形式 A 并赞成手术时，它是 1。若他或她赞成射线疗法时，它是 0。类似地，票上的第二个数是对形式 B 反应的代号。

随机不放回地从盒中抽取 80 次，并观察到对形式 A 的反应[A|]。对实验中形式 A 的反应如同第一批 80 个抽得数。盒中留下的 87 张票是第二个样本。对于这第二个样本，观察到对形式 B 的反应[|B]。对形式 B 的反应如同这第二批的 0 和 1。表 1 中的 50% 与第一个样本中 1 的百分率相同。表 1 中的 84% 与第二个样本中 1 的百分率相同。

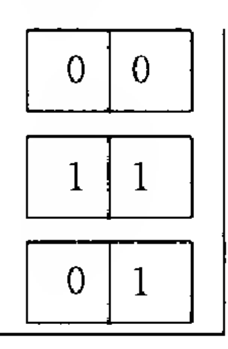
现在可以按照模型设立原假设。由于两种形式传达了同样的信息，经济学家认为医生对这两种形式的反应应该相同，因此票上的两个数是相同的(图 2)。盒子模型可用来证明我们的方法给出了 SE 的一个保守的估计[12]。

图 2　实验的原假设：根据具有相同的信息的形式 A 或形式 B，在手术和射线疗法之间作出抉择。

一个赞成射线疗法的医生

一个赞成手术疗法的医生

一种不可能的组合



这项研究中所使用的实验设计似乎有点不直截了当：为什么不把两种形式的问卷挨次分给所有的医生呢？原因很简单。这将导致提问题的一种方式会影响第二种方式提问题时的答案[13]。

习题 D

1. 对另一组实验对象：Harvard（哈佛）和 Stanford（史坦福）的 MBA 学生，重复了课文中所述研究。
   (a) 一个 MBA 学生，若给他形式 A 时会赞成射线疗法，但是给他形式 B 时会赞成手术疗法。填写他的票子 | A | B |。
   (b) 另一个 MBA 学生有票子 | 1 | 0 |。她对形式 A 的反应是什么？对形式 B 呢？
   (c) 下面列出的三张票中哪一张与原假设一致？
   (i) | 1 | 0 |　　(ii) | 0 | 0 |　　(iii) | 0 | 1 |
   (d) 结果解得如下。

	形式 A	形式 B
赞成射线疗法	84	17
赞成手术疗法	112	84

   对两种形式的反应的差异能用机会来解释吗？（提示：开始先求出有多少学生得到形式 A；他们中赞成射线疗法的百分率是多少？然后对形式 B 如法炮制。）
2. 在脊髓灰质炎疫苗现场试验中，400 000 名儿童参与一项随机化对照双盲实验：他们中的半数被随机指派在疫苗组，另一半在安慰剂组[14]。在疫苗组中，有 57 例小儿麻痹症，对比安慰剂组的 142 例。这个差归因于机会吗？若不是，怎样解释它？
3. 20 世纪 60 年代，提出了一种"负收入税"，对低收入进行补助而不是向他们收税。这项福利计划会造成受益人不工作吗？在 New Jersey（新泽西州）的三个城市做了一项实验来寻求答案。指标母体由这些城市的 10 000 个低收入家庭组成。从这些家庭中随机选出 400 个家庭作为对照组。另外随机选出 225 个为处理组——实施负收入税。对所有 625 个家庭跟踪 3 年[15]。
   (a) 对照组家庭在 3 年期间内平均受雇工作 7 000 小时，它们的 SD 是 3 900小时。处理组家庭则平均 6 200 小时。它们的 SD 是 3 400 小时。

两个平均数之间的差是统计显著的吗？你得出什么结论？

(b)在对照组中，88%的户主受雇，对比处理组的82%。差是统计显著的吗？你得出什么结论？

4. 在某个大的学区里，对250名中学生的简单随机样本进行了一次地理考试。有一道题涉及一张欧洲的简略地图，图中各国仅以编号区分。要求学生辨认出大不列颠和法国。考试结果，65.8%能找出法国，相比较，70.2%找出大不列颠[16]。差是统计显著的吗？还是这能由所给的信息来确定？

5. 婴儿出生后的前几个月以母乳喂养被认为对婴儿的健康比人工喂养好。根据一些观察研究，医院的保育室不供应奶瓶增加了母亲出院后继续母乳喂养的可能性。因此，提出了不供应补助食品的建议。

1982年Montreal（蒙特利尔），Royal Victoria（皇家维多利亚）医院的K. Gray－Donald，M. S. Kramer和他的同事做了一次对照实验[17]。有两间保育室，在"传统"的一间里，象往常一样给新生婴儿供应补助进食——凌晨2点一瓶，以及喂母乳后婴儿仍感饥饿的任何时候。在进行实验的一间里，在凌晨2点时唤醒母亲，让她们给自己的婴儿喂乳；并劝阻喂以补助进食。

在为期4个月的实验期内，393位母亲以及她们的婴儿被随机指派在传统保育室，388位到实验保育室。通常在医院的逗留期是4天，出院后进行9个星期跟踪。

(a)在9星期结束时，指派在传统保育室的母亲中有54.7%仍继续以母乳喂养她们的婴儿，对比实验保育室的54.1%。这个差异是统计显著的吗？你得出什么结论？

(b)若真的轮到母亲们决定母乳喂养还是人工喂养，她们的决定因处理而改变吗？为了回答这一问题，调研人员查看了两间保育室里人工喂养的用量，以日毫升数（毫升/日）表示。在传统保育室里，每个婴儿平均36.6毫升/日，SD是44.3；在实验保育室里，统计数据是15.7和43.6。你得出什么结论？

(c)两间保育室里的不同的待遇以任何方式影响婴儿吗？为了回答这个问题，调研员查看了每个婴儿住院期间体重的减少量，以出生体重的百分率表示。在传统保育室里，平均减少5.1%，SD是2.0%。在实验保育室，平均数是6.0%，SD是2.0%。你得出什么结论？（可能感到惊奇，但是，绝大多数新生婴儿在生命的最初几天会减少一点体重。）

(d)随机化成功吗？为了查明，调研人员（在其它量之中）考虑了出生体重。

在传统保育室，平均 3 486 克，SD 是 438 克；在实验保育室，平均数是 3 459 克，SD 是 434 克。你得出什么结论？

这些习题的答案在第 727—729 页上。

5. 复习题

复习题可能包含前面几章的内容。

1. 从一盒标有数字的票中随机放回地抽取 500 次；276 次是正数。某人告诉你说盒中 50% 的票子标的是正数。你相信吗？回答是或否，并解释之。
2. 从盒 A 中随机放回地抽取 100 次，另从盒 B 中随机放回地抽取 250 次。
   (a) 从盒 A 中抽得的 50 个数是正的，与从盒 B 抽得的 131 个数是正的相比：50.0% 对 52.4%。这个差是实际存在的，还是归因于机会？
   (b) 盒 A 的抽得数平均为 27.4，它们的 SD 是 15.3；盒 B 的抽得数平均为 32.3，它们的 SD 是 16.1。两个均值间的差统计显著吗？
3. Gallup 民意测验询问回答者，他们对美国各种机构有多大信任[18]。1987 年，在回答者中仅 52% 对最高法院有“巨大或相当多”的信任，相比于 1979 年的 60%。这个差是实际存在的吗？或你能根据所给信息回答吗？你可以假设这些结果是基于每年各 1 000 人的独立简单随机样本。简短地讨论之。
4. Gallup 民意测验询问回答者，他们对美国各种机构有多大信任。1987 年，31% 的回答者对报纸有“巨大或相当多”的信任，相比于对电视的 28%[18]。这个差是实际存在的吗？或你能根据所给信息回答吗？简短地讨论之。你可以假设这些结果是基于 1 000 人的简单随机样本。
5. 一项实验包括 British Columbia 大学的 383 名学生：随机选出 200 人得到条款 A，他们中有 92 人回答“是”；其余 183 人得到

条款 B,第二组中的 161 人回答“是”[19]。

条款 A)设想你决定观看某场演出并付了每张票价为 10 美元的入场费。当你进入戏院时,发现丢失了票。坐位是没有编号的,票子又没能重新找到。你会再付 10 美元另买一张票吗?

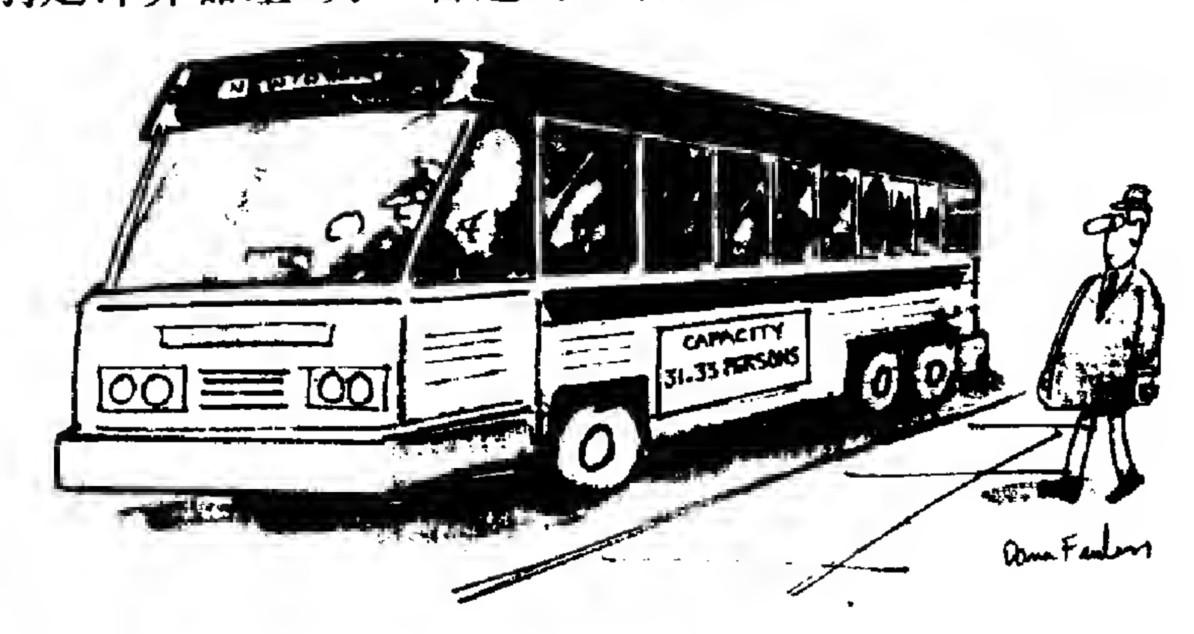
条款 B)设想你决定观看某场演出,票价是每张 10 美元。当你来到戏院时,你发现你丢失了一张 10 美元的钞票。你仍然会付 10 美元买票吗?

根据经济学理论的观点,两个条款提出了相同的事实而要求相同的回答;它们间的任何差异必定归于因于机会。从心理学的观点,问卷的结构可以预期会影响回答。数据说明了什么?

6. 实施一项实验以了解计算器是否帮助学生解数学文字题[20]。实验对象是某学区的一组 500 人的 13 岁儿童。所有实验对象都解下面的题目。从他(她)们中随机选出一半,并允许使用计算器,其余用铅笔和纸解题。在计算器组,18 名学生得出正确答案;在铅笔和纸组,是 59 名。这个差能用机会解释吗? 你得出什么结论?

问题:一辆军车可载 36 名士兵,若要将 1 128 名士兵载赴训练场,需要几辆军车?

注:1 128/36=31.33 故需要 32 辆,但是一般的回答是 31.33,特别是计算器组,另一普遍的回答是 31。



7. 当罪犯从牢里释放时，他们没有钱，且有很高的“再犯罪”率：释放的罪犯重又犯罪并再次被捕。对获释罪犯在他们获释后的第一个月发放收益资助能减少再犯罪吗？劳工部搞了一次随机对照实验以寻找答案[21]。实验对选出的一组自 Texas 和 Georgia 的某些监狱释放的罪犯进行。收益资助全按失业保险金方式发放。有一个不发补助的对照组和四个不同的处理组(所付资助略为不同)。

习题是关于 Georgia 州的结果的，并将四个处理组合并为一个。假设罪犯是随机指派到处理组或对照组。

(a)592 名罪犯分在处理组，他们中 49.3%释放后一年内再次被捕。154 名分在对照组，他们中的 48.4%释放后的一年内再次被捕。收益资助会减少再犯罪吗？回答是或否，并作简短解释。

(b)在他们从监狱里释放后的头一年中，分在处理组的平均受雇 16.8 周；SD 是 15.9 周。对那些指派在对照组的，平均受雇 24.3 周。SD 是 17.3 周。收益资助减少获释罪犯参加工作的时数吗？回答是或否，并作简短解释。

8. 人们可以预言他们表现将如何好吗？预言影响行为吗？一项实验比较对于“预言——要求”的反应与“只要求”反应[22]。在预言——要求组中，先要求实验对象预测他们是否同意做某些志愿者工作，然后要求他们做该工作。在“只要求”组中，要求实验对象做该工作；在此之前不要求他们预测。在下列 a—c 三个部分中，两样本 Z—检验也许合用也许不合用，若它是可行的，做 Z—检验。若不行，为什么不行呢？

(a)随机选出 Indiana 州 Bloomington 镇的 46 个居民为“预言——要求”处理组。打电话要求他们表态，“如果通过电话向他们提出要求时，他们是否同意花 3 个小时为美国癌症协会募集捐款，”46 人中的 22 人称他们同意。该镇随机选出的另外 46 人为“只要求”处理组：要求他们花 3 小时为美国癌

症协会募捐款。46 人中仅 2 人同意做这事。22/46 与 2/46 间的差能归因于机会吗？你得出什么结论？

(b)三天后，再次给预言——要求组打电话，要求他们花 3 个小时为美国癌症协会募集捐款。46 人中的 14 人同意做这事。14/46 与 2/46 之间的差能归因于机会吗？你得出什么结论？

(c)22/46 与 14/46 之间的差能归因于机会吗？你得出什么结论？

9. 20 世纪 70 年代，多风险因素防治试验对降低冠性心脏病的三种风险因素—血清胆固醇，血压和抽烟的防治措施的作用进行检测。实验对象是12 866名 35 岁到 57 岁处于心脏病高风险的男人。6 428 名随机地指派去防治组而 6 438 名去对照组。防治措施包括劝阻抽烟，节制饮食以及某些情况下的治疗以降低血压。实验对象至少受跟踪 6 年[23]。

(a)进入研究时，防治组的舒张压平均是 91.0mmHg；他们的 SD 是 7.6mmHg。对照组的统计数字是 90.9 和 7.7，你得出什么结论？（血压以水银柱毫米，或 mmHg，计量。）

(b)6 年后，防治组的舒张压平均是 80.5mmHg，他们的 SD 是 7.9mmHg。对照组的统计数字是 83.6 和 9.2。你得出什么结论？

(c)进入研究时，防治组的血清胆固醇的水平平均为 253.8mg/dl；他们的 SD 是 36.4mg/dl。对于对照组统计数字是 253.5 和 36.8。你得出什么结论？（mg/dl 是每分升的毫克数。）

(d)6 年后，防治组的血清胆固醇水平平均是 235.5mg/dl；他们的 SD 是 38.3mg/dl。对照组的统计数字是 240.3 和 39.9。你得出什么结论？

(e)进入研究时，防治组的 59.3%对象抽烟，对比对照组有 59.0%的人抽烟。你得出什么结论？

(f)6 年后，抽烟者的百分率是：防治组是 32.3%，对照组是 45.6%。你得出什么结论？

(g)在处理组中，211 人 6 年后死亡，对比于对照组有 219 人死亡。你得出什么结论？

10. 某科研工作者想知道社会工作领域的期刊编辑是否有偏向。他把研究有关"把一个患气喘病的儿童从他的父母身边隔离开，以便减轻常常由心理引起的身心疾病的症状"的论文整理成两种不同的说法。在一种说法中，对隔离持肯定效果；在另一种说法中，则为否定[24]。论文投寄给 107 个期刊；随机选出的 53 个期刊收到肯定说法，其余 54 个收到否定说法。结果如下。

	肯定	否定
接受	28	8
拒绝	25	46

上表的第一列表明 28 个收到肯定说法的期刊接受发表，25 个拒绝；第二列给出收到否定说法的结果。你得出什么结论？

11. 某调研人员想证明头胎儿童智商测验的得分高于第二胎的儿童。在某学区，他找到了两个小孩都在小学就读的 400 个双子女家庭。他给这些小孩做 WISC 词汇量等级测验（如第 553 页上习题 3 所述），结果如下：

- 400 个头胎儿童平均得 29 分，他们的 SD 是 10 分。
- 400 个第二胎儿童平均得 28 分，他们的 SD 是 10 分。

得分是对不同年龄收集的。然后他做两样本 Z—检验：

头胎平均数的 SE≈0.5

第二胎平均数的 SE≈0.5

差的 SE $\approx\sqrt{0.5^2+0.5^2}$

≈ 0.7

P≈ ≈8%

1.4

$Z=(29-28)/0.7\approx 1.4$

对统计检验的使用作简短地评论。

6. 小结

1. 两个独立量的差的标准误差是$\sqrt{a^2+b^2}$，其中

- a 是第一个量的 SE
- b 是第二个量的 SE

这公式不适用于两个相依量的差。

2. 假设两个相互独立且适当大的简单随机样本取自两个分开的盒子。原假设称这两个盒子有相同的平均数，故两个样本平均数之间的期望差为 0。合适的检验统计量为

$$Z=\frac{\text{观察差}-\text{期望差}}{\text{差的 SE}}$$

基于这个统计量的检验称为两样本 Z—— 检验。

3. 通过在盒子中放入 0 和 1，这个方法，可以处理包含分类和计数的情况。

4. 这个方法也可用来比较一个实验中处理组与对照组的平均数或百分率。假设有一盒票。每张票上有两个数：一个表示对处理 A 的反应；另一个是对处理 B 的；仅一个数可被观察到。从盒中随机不放回地抽取一些票子，并观察到对处理 A 的反应。然后，从剩下的票中随机不放回地抽取第二个样本。在第二个样本中，观察到对处理 B 的反应。两个样本平均数之间的差的 SE 可以保守地估计如下：

(i) 在随机放回抽取的基础上计算两个样本平均数各自的 SE。

(ii)象两个样本是相互独立那样汇总它们的 SE。

28

χ^2-检验

不要问它的含意，宁可问如何使用它。

——L. Wittgenstein(1889—1951)

1. 引言

它对事实的拟合有多好？对任何机会模型，这个问题迟早必须询问。在许多情况下，它可以用 χ^2-检验（1900 年由 Karl Pearson 创立）[①]解决　。χ 是希腊字母，象英语单词“kite”中的“ki”那样读；故 χ^2 读作“ki—方”。第 26 章第 5 节讲述了如何检验心灵学试验的机会模型。在那里，每个猜测都被分类到两个类——正确或不正确之一。根据模型，一个猜测有 1/4 的机会是正确的，故正确猜测的次数等于取自盒子

0	0	0	1

的抽得数之和。在这种情况下，Z—检验是适用的。因为只涉及两个类。在许多场合，模型将涉及多于两个的类，Z—检验可能不再适用。

譬如，假设掷一颗骰子。你想知道骰子是否公平——每一面都有六分之一的机会出现。现在，每掷一次可以归到 6 个类——出现

⚀，或⚁，或⚂，或⚃，或⚄，或⚅。下面的例子说明 χ^2-检验如何可以用来解决多类的问题。

例 1. 某赌徒被指责使用一颗灌过铅的骰子，但他辩称自己无罪。一组记录保存了最近的 60 个掷得数（表 1）。如何解释这批数据存在着分歧，为此请了一位统计学家。

表 1　一只可能灌过铅的骰子的 60 个掷得数

4	3	3	1	2	3	4	6	5	6
2	4	1	3	3	5	3	4	3	4
3	3	4	5	4	5	6	4	5	1
6	4	4	2	3	3	2	4	4	5
6	3	6	2	4	6	4	6	3	2
5	4	6	3	3	3	5	3	1	4

讨论. 如果这个赌徒是无罪的，表 1 中的 60 个数如同从盒子

| 1 | 2 | 3 | 4 | 5 | 6 |

中抽取 60 次（随机有放回）的结果。根据这个模型，每个数应该出现 10 次左右：期望频数为 10。要想查明数据与期望相比较如何，你必须数一下每个数实际出现几次。观察频数列在表 2 中。作为算术运算的核对，将各频数列相加：和必须等于 60，因为它是表 1 所列元素的总数。

表 2　表 1 数据的观察和期望频数

呈现值	观察频数	期望频数
1	4	10
2	6	10
3	17	10
4	16	10
5	8	10
6	9	10
和	60	60

正如该表所指出的，有太多的 3。3 的次数的 SE 为 $\sqrt{60}\times$

$\sqrt{1/6\times5/6}\approx2.9$，故观察到的次数约高出期望次数 2.4 个 SE 左右。但还未击中该赌徒要害。统计学家不会建议一次取该表的一行：

- 表中有好几行看来可疑。譬如，在表 2 里也有过多的 4。
- 另一方面，对于表里的许多行，有高概率表明它们中至少有一行看来可疑——即使模型是正确的。它就象玩俄罗斯轮盘赌。如果你不断玩下去，迟早你要输。

对表 2 中的每一行，观察和期望频数之间存在着差异。想法是将所有这些差汇总列在数据与模型的期望之间距离的一个整体的度量中去，χ^2—统计量对每一个差取平方，然后除以相应的期望频数，并取它们的和：

$$\chi^2=\frac{(\text{观察频数}-\text{期望频数})^2}{\text{期望频数}}\text{的和}$$

表中的每一行在和式中都占有一项。

乍一看，这公式也许看来太武断而不值得去理解。然而，由于它的某种非常方便的特性，这将在稍后指出，每位统计学家都使用它。现在，请容忍这个公式。

对于表 2，χ^2—统计量为

$$\frac{(4-10)^2}{10}+\frac{(6-10)^2}{10}+\frac{(17-10)^2}{10}+\frac{(16-10)^2}{10}+\frac{(8-10)^2}{10}+\frac{(9-10)^2}{10}=\frac{142}{10}=14.2$$

当一个观察频数远离期望频数时，和中的对应项就大；当两者接近时，这一项就小。大的 χ^2 值表明观察频数远离期望频数。小的 χ^2 值含意正好相反：观察频数接近期望频数。因此，χ^2 给出了观察频数与期望频数之间距离的一种度量②。

当然，即使表 1 中的数据由掷一颗公平的骰子 60 次产生，χ^2 仍有可能得出为 14.2 或更大：机会变异成了辩护律师。这似乎可能吗？要想查清楚，我们需要知道掷一颗公平的骰子 60 次，并用观察到的频数计算 χ^2，其值是 14.2 或更大的机会。

为什么要"或更大"？14.2 这个值可能成为推翻模型的证据，因为它太大，意思是观察频数离期望太远。如果真的如此，那么大于 14.2 的值将会是推翻模型的更强有力的证据。模型将产生这么强有力的证据来推翻它自己的机会有多大呢？为求得这个机会，我们计算一下获得 14.2 或更大的 χ^2—统计量值的机会。



计算这个机会看起来象是件令人生畏的工作，但在计算机上只需要几分钟，答案是 1.4%。若骰子是公平的，则它只有 1.4% 的机会产生一个象所观察到的那样大（或更大）的 χ^2—统计量值。至此，统计学家完成了他的工作。事情看来对该赌徒不太妙。

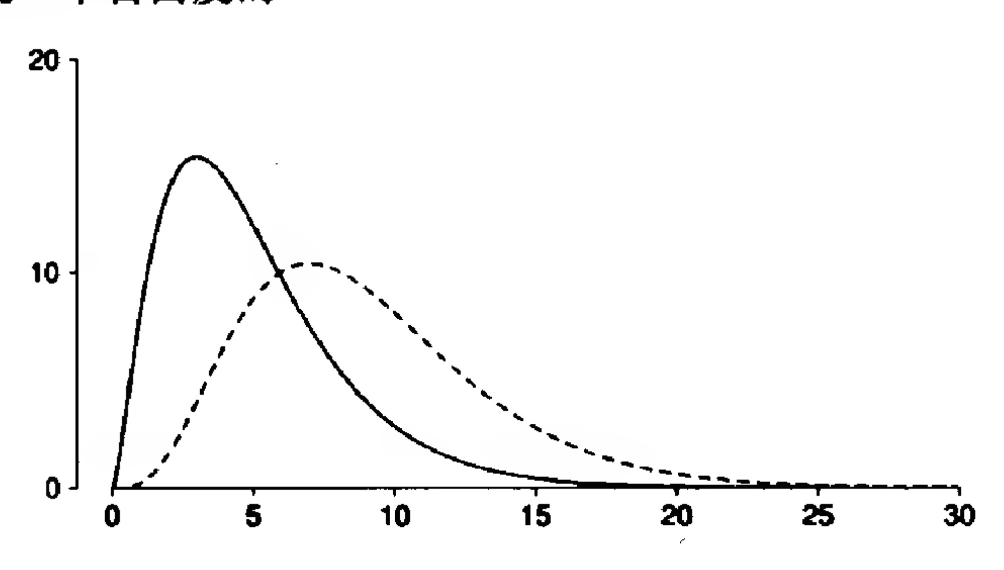
象在第 26 章中那样，1.4% 称为"观察的显著水平"并记为 P。在 Pearson 的年代，还没有计算机来做这工作。因此，他提出了一种用手算近似求 P 的方法，这涉及一条新的曲线，称为 χ^2—曲线。确切地说，对每一个自由度数目有一条曲线[3]。在例 1 中，模型是完全指定的：没有参数要由数据估计，因为模型告诉了你在盒子里的是什么。于是

$$\text{自由度} = \chi^2 \text{ 中的项数} - 1$$

（在 χ^2 公式中的所有差加起来必定等于 0；这约束消去了一个自

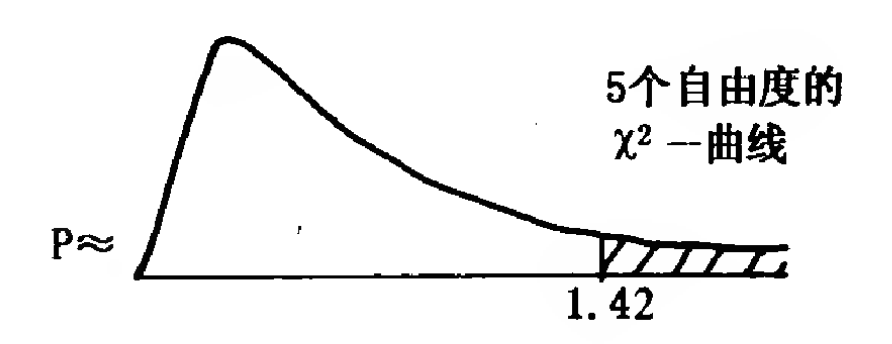
由度：见第 26 章第 6 节。）对于骰子此例，有 6－1＝5 个自由度。自由度为 5 和 9 的曲线在图 1 中给出。

图 1　自由度为 5 和 9 的 χ^2—曲线。这些曲线有长的右尾。当自由度增大时，曲线变扁平并向右偏移。（实曲线是 5 个自由度的；虚曲线是 9 个自由度的。）



对于 χ^2—检验，P 近似地等于 χ^2—统计量观察值以右，适当自由度的 χ^2—曲线下面的面积。当模型是完全指定时（没有参数需要估计）， 自由度＝χ^2 中的项数－1

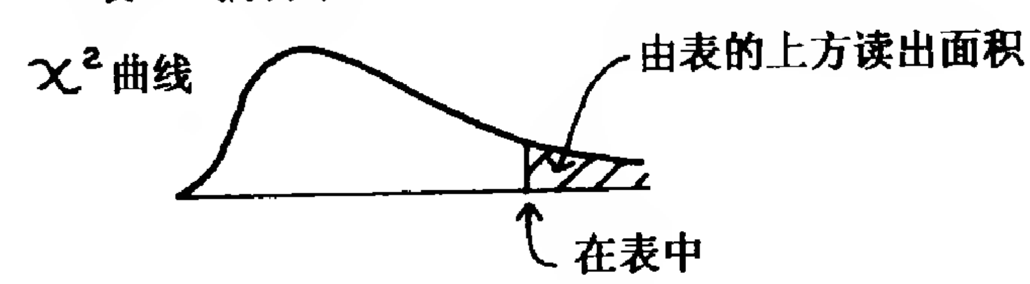
对于例 1



该面积可由表或统计计算器求得。原则上，每一条曲线有一张表，但这样使用太不方便，因此采用了不同的安排方式，如表 3 所示（摘自第 737 页上较大的一张表）。各种面积，按百分数单位，横贯地排列在表中最上面的一行里；各种自由度则由上而下罗列在表的左边。譬如，查 5％的一列和 5 个自由度的一行时。在表的中心腹部是元素 11.07，意思是 5 个自由度曲线下 11.07 右面的面

积是 5%。5 个自由度曲线下 14.2 右面的面积在表中找不到，但它介于 5%(11.07 右面的面积)和 1%(15.09 右面的面积)之间。有理由猜测这面积仅略大于 1%。

表 3　摘自第 735 页上一张较大 χ^2 表的简表



自由度	90%	50%	10%	5%	1%
1	0.016	0.46	2.71	3.84	6.64
2	0.21	1.39	4.60	5.99	9.21
3	0.58	2.37	6.25	7.82	11.34
4	1.06	3.36	7.78	9.49	13.28
5	1.61	4.35	9.24	11.07	15.09
6	2.20	5.35	10.65	12.59	16.81
7	2.83	6.35	12.02	14.07	18.48
8	3.49	7.34	13.36	15.51	20.09
9	4.17	8.34	14.68	16.92	21.67
10	4.86	9.34	15.99	18.31	23.21

Pearson 的近似有多好？图 2 给出了掷一颗公平骰子 60 次的 χ^2—统计量的概率直方图。同时也绘出一条自由度为 5 的 χ^2 曲线。直方图比曲线崎岖不平得多，但还是较好地拟合于它。特别地，直方图下任一给定值右面的面积接近于曲线下面相应的面积。这些尾部面积之比画在图 2 的底部。回到例 1，直方图下 14.2 右面的面积给出了 P 的确切值。它是 1.4382%。曲线下 14.2 右面的面积，给出了 P 的 Pearson 近似值，它是 1.4388%；近似程度不错④。

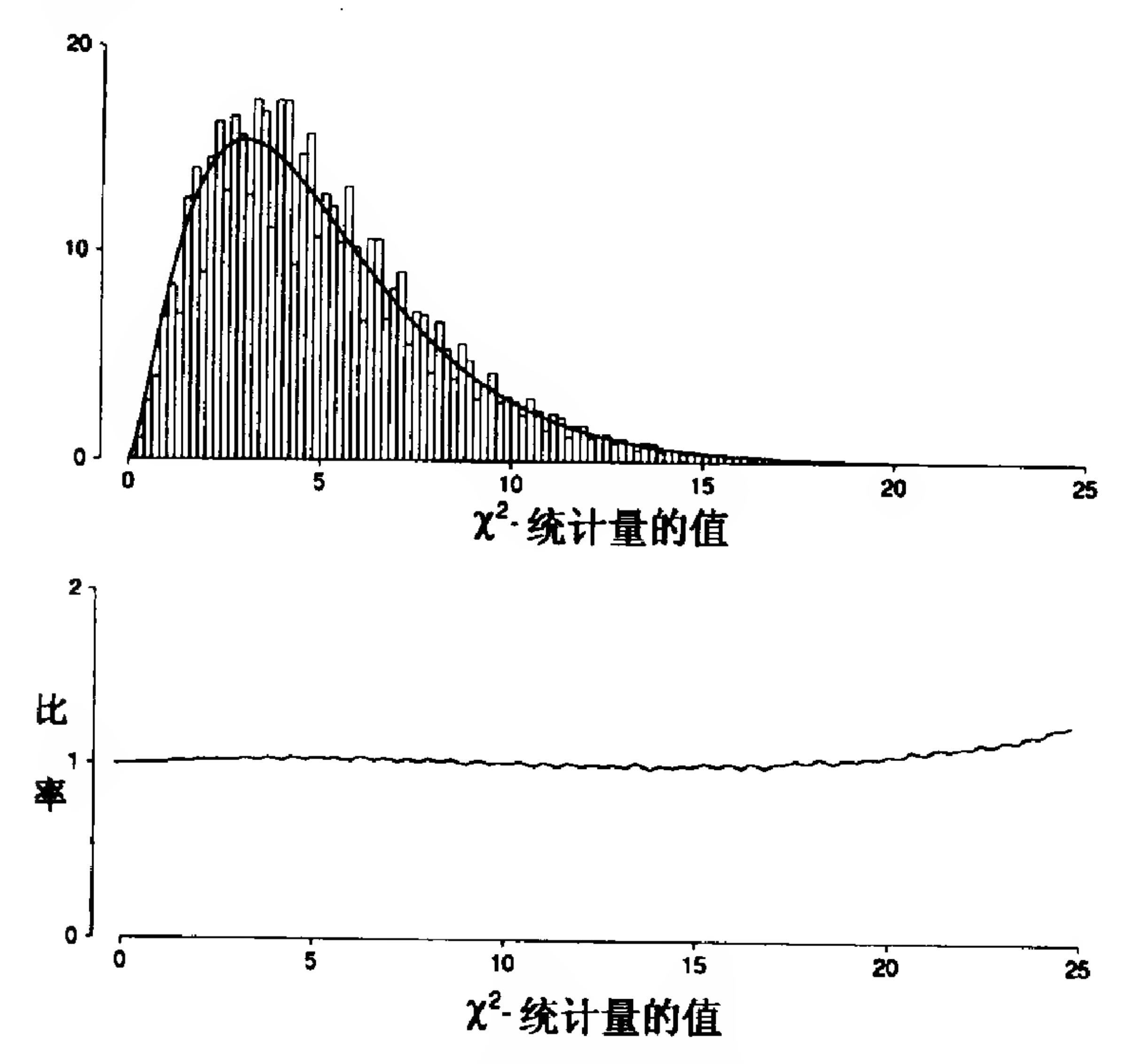
约略的估计，当表中每一行的期望频数是 5 或更大时，近似值是可信的。在表 2 中，各期望频数是 10，近似程度非常出色。另一方面，对于取自盒子

[1] [2] [3] 96 张 [4]

的 100 次抽得数，近似程度将不会有这么好。在这情况下，[1] 的期

望次数仅为 1；[2] 与 [3] 的也一样。期望频数小到使近似值不可信。

图 2 Pearson 近似值。上面一张图给出了掷一颗公平骰子 60 次的 χ^2—统计量的概率直方图，与(5 个自由度的) χ^2 曲线比较。下面的一张图表示了尾部面积之比。例如，在水平轴上取 14.2，直方图之下，14.2 右面的面积是 1.4382%，曲线之下相应的面积是 1.4388%。比值 1.4382/1.4388≈0.9996。描绘在 14.2 的上方。其余比值以相同方式描绘。



相对于 Z—检验，什么时候应该使用 χ^2—检验？若关系到盒子的组成百分率时，用 χ^2—检验。若仅涉及盒子的平均数，用 Z—检验。譬如，假设你从一盒编有数值 1 至 6 的票子中有放回地抽取；盒中不同类票的百分率未知。为了检验每个数值占票子的$16\frac{2}{3}\%$的假设，则用 χ^2—检验。基本上，满足这个假设的盒子只有一只：

1	2	3	4	5	6

另一方面，为了检验盒子的均值为 3.5 这一假设，则用 Z—检验。当然，除了 1 2 3 4 5 6 之外还有许多盒子其平均数为 3.5，譬如，

1 2 3 3 4 4 5 6 或 1 1 2 3 4 5 6 6

总结一下：

- χ^2—检验回答数据是否与从一只所装内容已知的盒子中随机抽取的结果一致。
- Z—检验回答数据是否与从一只平均数给定的盒子中随机抽取的结果一致[5]。

本节的余下部分讲述 χ^2 如何用于某种幸运轮盘[6]。California 州彩票的某些中奖者被选出参予称为"大转盘(The Big Spin)"的电视游戏。每个参赛者转一只重型铝制轮盘，其上有 100 只标号 1 至 100 的槽。一只硬质橡皮球在轮盘里蹦跳着，最后落入这个或那个槽中，并决定该奖给参赛者的奖金数目。

成百万的美元在赌注中，因此轮盘必须十分仔细地检查。州彩票委员会的统计顾问 Don Ylvisaker 请操作人员将轮盘转动 800 次，并记下球落入各槽的次数。然后他作了一次有关观察频数对期望频数的 χ^2-检验。结果 χ^2-统计量的值是 119。有 $100-1=99$ 个自由度，$P\approx 8\%$。这似乎勉强合格。

在试转中，69 号槽出现次数最多，19 号最少。在盘上这两个数彼此相对。于是将轮盘更仔细地检查一下：在背面找到一只金属重物附在 69 号槽附近的边上。显然，这样做是为了平衡轮盘，就象平衡一只汽车轮胎一样。

重物被去掉，轮盘重新平衡，并再次检测。起先转出的 400 个数看不出特别地随机，但自那以后情况有改善。实际上，约转到 400 次时，由于轮盘发出扎扎的怪声，操作人员加了点油。轮盘被接纳使用，且运转得很好。(定期地加油。)



“不！你不能给轮盘加油。”

2. χ^2—检验的构造

第一节叙述了 χ^2-检验。组成成分是些什么？

(a)基本数据。它由一定数目的观察值组成，通常记作N。对于骰子，N是60，基本数据列在表1中。对于幸运轮盘，N是800；基本数据是试转产生的800个数。

(b)机会模型。至今，本章只考虑了一种机会模型。有一盒票子，其内容给定；从盒中抽取是随机放回的：根据模型，数据就如同抽得数。对于骰子，盒子是

1	2	3	4	5	6

对于幸运轮盘，盒子里有100张票，编号从1到100。

(c)频数表[⑦]。对每一个值，观察频数由基本数据计数获得；期望频数由N和机会模型获得。表2记录了骰子的观察和期望频数。轮盘的频数表将有100行；这里从略。

(d)χ^2-统计量。它按公式计算。对于骰子，χ^2-统计量的值是14.2；对于轮盘，χ^2-统计量的值是119。

(e)自由度。它等于 χ^2 的和式中所含项数减 1(当盒子的内容给定时,如本节的情况。)对于骰子,有 5 个自由度,对幸运轮盘,有 99 个。

(f)观察的显著水平。它是适当的自由度 χ^2-曲线下面 χ^2-统计量值右面的面积近似。对于骰子,P≈1.4%;对于轮盘,P≈8%。

专门术语有点复杂,因为每个都以“χ^2”起头:

- χ^2-检验包含步骤(a—f);
- χ^2-统计量根据你每次做检验时的数据计算;
- 两条 χ^2-曲线在图 1 中给出;
- χ^2-表基于 χ^2-曲线,用来查 P 值。

无论在盒子里的东西是什么,只要 N 足够大,同样的 χ^2-曲线和表可用来近似求 P。这结果证明公式是合理的。对于其它检验统计量,每只盒子将需要一条新的曲线。

习题 A

1. 求 5 个自由度的 χ^2-曲线下面,及位于下值右面的面积。

   (a)1.61　　(b)9.24　　(c)15.09

2. 求 9 个自由度的 χ^2-曲线下 15.09 右面的面积。
3. 假设表 2 中的观察频数结果如下面表 4A 中所示。试计算 χ^2 的值,自由度和 P。能推断出什么?(在后面的习题中,这将简称为“做 χ^2-检验”。)
4. 假设表 2 中的观察频数结果如下面表 4B 中所示。试做 χ^2-检验。
5. 假设表 1 有 600 项而不是 60 项,其观察频数如表 4C 中所示。试对原假设即观察值由掷一颗骰子 600 次所产生做 χ^2-检验。
6. 假设表 1 中有 60 000 个元素,其观察频数如表 4D 中所示。

   (a)计算每个值显示次数的百分数。

   (b)骰子是公正的吗?

   (c)对原假设,做 χ^2—检验:数据是掷一只骰子 60 000 次所产生。

表 4A		表 4B		表 4C		表 4D	
值	观察频数	值	观察频数	值	观察频数	值	观察频数
1	5	1	9	1	90	1	10 287
2	7	2	11	2	110	2	10 056
3	17	3	10	3	100	3	9 708
4	16	4	8	4	80	4	10 080
5	8	5	12	5	120	5	9 935
6	7	6	10	6	100	6	9 934

7. 某人希望检验某骰子是否公正，并将它掷了 600 次。每掷一次，他只记录结果奇数还是偶数，以及是大(4,5,6)还是小(1,2,3)。观察频数结果如下：

	大	小
偶数	183	113
奇数	88	216

做 χ^2—检验以查看骰子是否公正。

8. 某项有关 California 州，Alameda 县的大陪审团的研究，比较了陪审员们与全体人口的人口统计特征，目的是查看陪审团的陪审员名单是否具代表性[8]。这里是一份关于年龄方面的结果。(仅考虑 21 岁和以上的人；县的年龄分布由公共卫生部的数据而得知。)

年龄	全县的百分数	陪审员人数
21 到 40	42	5
41 到 50	23	9
51 到 60	16	19
61 及以上	19	33
	—	—
总　数	100	66

这 66 名陪审员是从 Alameda 县的(21 岁及以上)人口总体中随机选　取的吗？

9. California 州彩票委员会检测的另一套设备有一组编号从 0 到 9 的 10 只乒乓球。这些球用空气喷射方法在一只玻璃滚筒里搅混，并随机强行逼出一只。在按下述方式进行的试验中，搅混机看来运转良好，但是有些球组也许表现不好。在每一轮试验中，机器从滚筒里做 120 次有放回抽取。

(a)假设一切正常。从滚筒抽取的 120 次中，每只球可期望被抽到____次。

(b)若球 7 出现 29 次，你会得出什么结论？为什么？

(c)下面的表给出了检测 4 组球，以及重复检测它们中的 2 组的结果。A 组和 D 组被重复检测；B 组被拒绝；C 组被接纳使用。为什么？（读表：对球组 A，球 0 被抽出 13 次；球 1 被抽出 11 次；等等）

(d)重复检测后，对 A 组球和 D 组球你将做些什么？简短地解释。（该表：对 A 组球，在重复检测一列，球 0 被抽出 19 次，球 1 被抽出 9 次；等等。）

频　　数

球号	A 组球		B 组球	C 组球	D 组球	
	检测	重复检测	检测	检测	检测	重复检测
0	13	19	22	12	16	8
1	11	9	8	10	7	15
2	16	10	7	14	12	22
3	11	12	8	10	14	11
4	5	7	19	11	15	15
5	12	15	20	10	5	8
6	12	19	10	20	10	17
7	19	10	11	12	21	9
8	5	12	6	12	11	8
9	16	7	9	9	9	7

10. 某人告诉你用下面的方法解前面一道习题：将各数转化成百分数；譬如 66 次中的 5 次是 7.6%；取观察和期望百分数之间的差；将差平方；再除以期望百分数；相加得出 χ^2。这对吗？

这些习题的答案在第 730—731 页上。

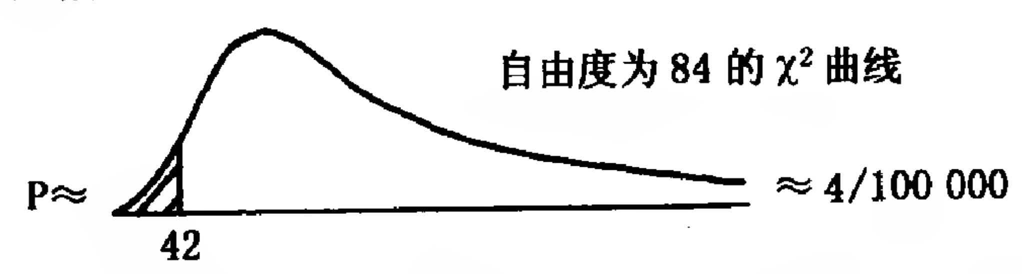
3. Fisher 如何使用 χ^2—检验。

Fisher 用 χ^2—检验证明 Mendel 的数据（第 25 章）是捏造的[9]。对 Mendel 的每一个实验，Fisher 计算其 χ^2—统计量。这些实验全部都相互独立，因为它们包含不同的作物组。Fisher 将结果合并。

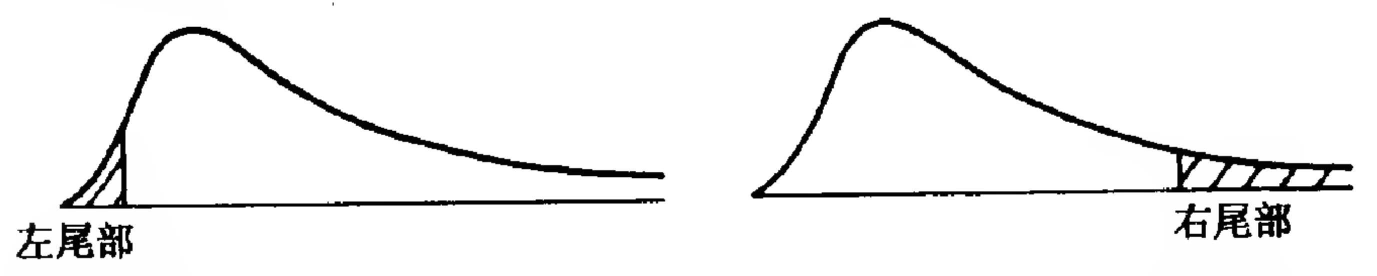
> 对于独立实验，可以通过将个别 χ^2-统计量相加来合并结果；自由度也相加。

譬如，一个试验给出 $\chi^2=5.8$，自由度是 5，另一个独立试验给

出 $\chi^2=3.1$，自由度是 2，二个合起来得到 $5.8+3.1=8.9$ 的合并 χ^2，其自由度是 $5+2=7$。对于 Mendel 的数据，Fisher 获得合并 χ^2—值是 42 不到一点，具自由度 84。42 左边，自由度为 84 的 χ^2—曲线下面的面积为 100 000 分之 4。观察与期望之间的一致好得使人不能相信。



在这一点看来涉及一个新的原则：P 被计算为左尾部面积，不是右边的一个。为什么？



原因在这里。Fisher 不是在检验 Mendel 的机会模型；他认为它是正确的。取而代之，他在比较两个假设——

- 原假设：Mendel 的数据是真实收集的。
- 备择假设：Mendel 的数据是捏造的，目的是使得记录的频数更接近于期望的频数。

小的 χ^2 值表示观察频数比起机会变异所能允许的更接近于期望频数，并赞成备择假设。由于是小的 χ^2 值拒绝原假设，P 必须计算为左尾部的面积。

习题 B

1. 设同一颗骰子用来产生习题 A 中第 3 和第 5 题的数据，先掷 60 次供习题 3，然后再掷 600 次供习题 5。你能合并这二个检验的结果吗？若行，如何进行？
2. 设同一颗骰子用来产生习题 A 中第 3 和第 5 题的数据，共掷 600 次，开头的 60 次用于习题 3；而习题 5 记录了所有 600 次结果。你能合并这两个检

验的结果吗？若行，如何进行？

3. 一项 Mendel 的繁殖试验结果如下[9]。做 χ^2—检验核实一下这些数据是否捏造。证据指出是哪一种情况？它是确凿的吗？

豌豆类型	观察数	期望数
光滑黄色	315	313
起皱黄色	101	104
光滑绿色	108	104
起皱绿色	32	35

这些习题的答案在第 731 页上。

4. 独立性检验

χ^2—统计量也可以用来检验独立性，这一点将在本节阐述。然而，讨论是相当技术性的。故可以跳过不读。方法将用一个例子简明地说明：习惯用那一只手和性别独立吗？更确切地说，取某总体——譬如，美国全体 25—34 岁的人。问题是在这个总体中，在男人中间“用手习惯”（习惯用右手，习惯用左手，左右手都善用）的分布是否不同于在女人中间的分布。

如果总体中的每一个男人和女人，他（她）们是习惯用右手，习惯用左手，或左右手都善用的数据是现成的，那么只要通过计算百分数，就可能直接解决这个问题。这种信息并不现成。不过，HANES（第 64 页）取了一个 2,237 名 25—34 岁的美国人的概率样本，他们对每个样本人的查看内容之一，是有关用手习惯。结果列在表 5 中。

表 5　按性别的用手习惯

	男　人	女　人
习惯用右手	934	1 070
习惯用左手	113	92
左右手都善用	20	8

这是一张 3×2 的表，因为它有 3 行和 2 列。一般地，在研究两个变量之间的关系时，若其中一个有 m 个值，另一个有 n 个值，则需要一张 m×n 的表。

根据表 5，约 88%的男人习惯用右手，比较于女人中有 91%。女人比男人稍许更可能习惯用右手。根据现代神经生理学，习惯用右手与大脑里左半球优势度相联系，理性官能支配感性官能[19]。样本表明女人比男人更具理性吗?另一种解释：习惯用右手是在全社会中赞许的，习惯用左手是在全社会中举止异常的。女人是否比男人在更大的压力下去遵循用手习惯的社会规范?

一个不那么戏剧性的解释：它只是机会变异。毕竟，即使总体中全体男人中间用手习惯的分布与女人的分布完全相同，样本中也会出现不同。单凭抽取中的运气也可能使样本中有过少的习惯用右手的男人，或过多的习惯用右手的女人。为确定观察差是实际存在的或者只是归因于机会变异，需要做统计检验；这正是 χ^2—检验起作用之处。如果你并不担心机会变异，你就用不着做显著性检验。

HANES 的抽样设计过于复杂而不能用 χ^2—检验方法来分析。（这问题是抽样中产生的，在抽样中 SE 的公式依赖于设计；第 427 页。）为了讲述这个技巧，我们将假装表 5 是基于一个从总体中随机不放回地选取的 2 237 人的简单随机样本。于是，数据的盒子模型可建立如下。对应总体(25—34 岁的美国人)中的每一个人，盒子里有一张票。这成百万张的票的每一张都按照下面方式之一标明，描述着对应的人：

?? 习惯用右手的男人	?? 习惯用右人的女人
?? 习惯用左手的男人	?? 习惯用左手的女人
?? 左右手都善用的男人	?? 左右手都善用的女人

成百万张的票

我们的抽样模型说表 5 中的诸数是按下面方式产生的，即从这巨大的盒子中随机不放回地抽取 2 237 张票，并一一数出 6 种

不同类型的每一种有多少张。盒子的百分数成份是未知的，因此模型中有 6 个参数。

现在我们可以按照盒子模型的方式设立原假设和备择假设。原假设称用手习惯与性别相互独立；总体中男人习惯用右手的百分数与相应女人的百分数相等；对习惯用左手和左右手都善用类同。根据这一假设，样本百分数中的差只是反映了机会变异。备择假设称具相依性：盒中全体男人中间用手习惯的分布不同于女人中的分布。根据这一备择假设，样本中的差反映了总体中的差。

为了做原假设的 χ^2—检验，我们必须将观察频数（表 5）与期望频数相比较。计算期望频数的技巧似乎有点复杂。故将在稍后解释。期望频数本身在表 6 中给出：它们是在独立性的原假设基础上算出的。

表 6　观察与期望频数

	观察的		期望的	
	男人	女人	男人	女人
习惯用右手	934	1 070	956	1 048
习惯用左手	113	92	98	107
左右手都善用	20	8	13	15

下一步是计算

$$\chi^2 = \frac{(\text{观察频数}-\text{期望频数})^2}{\text{期望频数}}\text{之和}$$

$$= \frac{(934-956)^2}{956} + \frac{(1070-1048)^2}{1048} + \frac{(113-98)^2}{98}$$

$$+ \frac{(92-107)^2}{107} + \frac{(20-13)^2}{13} + \frac{(8-15)^2}{15}$$

$$\approx 12$$

有多少个自由度？

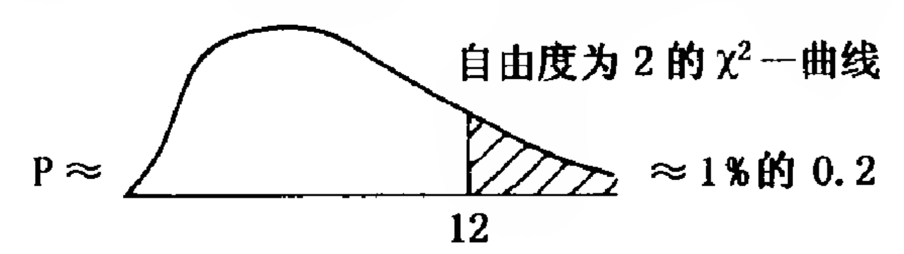
> 当检验一张 m×n 表（不具有有关概率的其它约束）中的独立性时，有(m－1)×(n－1)个自由度。

χ^2 的和式中有 6 项，但仅有(3－1)×(2－1)＝2 个自由度。要知道为什么，考虑下面这些差

－22	22
15	－15
7	－7

(计算第一个差的算术运算：934－956＝－22。)这些差按水平方向或竖直方向加起来都等于 0。因此，比如说如果你知道了－22 和 15，你可以算出所有其它几个数；只有两个差是自由变化的。

现在我们有了 χ^2—统计量和它的自由度，容易算出 P：



观察的显著水平 P 是自由度为 2 的 χ^2—曲线下 12 右面的面积，约为 1%的 0.2，原假设应予拒绝。样本中的观察差看来反映了总体中实际存在的差，而不是机会变异。这就是 χ^2—检验所宣称的。(更仔细的分析应该考虑样本设计，不过结论仍相同[11]。)

剩下来的问题就是计算表 6 中的期望频数，这得费点力气。你可以从计算表 5 的行和列的总数着手，如表 7 中所示。

表 7　行和列的总数

	男人	女人	总数
习惯用右手	934	1 070	2 004
习惯用左手	113	92	205
左右手都善用	20	8	28
总　　数	1 067	1 170	2 237

由表 7，样本中习惯用右手者的百分数是

$$\frac{2\ 004}{2\ 237}\times 100\%\approx 89.6\%$$

男人数是1 067。若用手习惯与性别相互独立，样本中习惯用右手的男人数应该是

1 067的89.6%≈956

表6中的其余期望频数可以用同样方法求出。“期望的”意指在独立性原假设上。

习题C

1.（假设的.）某城镇有约一百万合法选民，从中选取一个容量为10 000人的简单随机样本，以研究上届选举中性别与参予的关系。结果是：

	男人	女人
投票	2 792	3 591
没投票	1 486	2 131

原假设是性别和投票相互独立，做原假设的 χ^2—检验。

2. 下面的表是在Wyoming，1988年3月现场人口调查中25—29岁人口的交叉列表[12]。男人和女人的分布相同吗？你如何解释这些结果？你可以假设这些数据取自一个简单随机样本。

	男人	女人
从未结过婚	21	9
已　婚	20	39
丧偶/离婚/分居	7	7

这些习题的答案在731页上。

5. 复习题

复习题可能包含前面几章的内容。

1. 作为关于选择Alameda县大陪审团的研究的一部分，特将大陪审团陪审员的受教育水平与县的受教育水平分布相比较[13]：

受教育水平	县	陪审员人数
初等	28.4%	1
中等	48.5%	10
上过大学	11.9%	36
大学程度	11.2%	35
总数	100%	62

从该县取出的一个62人的简单随机样本能够给出其受教育

水平的分布与该县的分布如此地不同吗？选下面见解之一，并作解释：

(i) 这绝对不可能。

(ii) 这是可能的，但非常不象。

(iii)这是可能的，但不象——机会约在1%左右。

(iv) 这完全有可能——机会约在10%左右。

(v) 这几乎是肯定的。

2. 1988年3月的现场人口调查中，每一个答卷人被分类为受雇的，失业的，或劳动力以外的。California州35—44岁的男人的分类结果可以按婚姻状况交叉列成表如下[14]：

	已婚	丧偶、离婚或分居	从未结过婚
受雇的	638	133	102
失业的	27	8	6
不在劳动力中	35	12	20

譬如，已婚男人的91%受雇，相比较，从未结过婚男人有80%。这些差能解释成机会变异的结果吗？(你可以假设表上的结果来自一个简单随机样本。)

3. 某人声称他正在掷的是一对公平的骰子。为了检验他的声称，你让他掷这对骰子360次，并数出各种和数出现的次数。结果给出如下表。(为使你方便起见，一对公平骰子掷出每种和数的机会也同时给出。)你能和这个人玩掷骰子赌博吗？或者这些观察频数过份地接近期望频数吗？

和	机会	频数
2	1/36	11
3	2/36	18
4	3/36	33
5	4/36	41
6	5/36	47
7	6/36	61
8	5/36	52
9	4/36	43
10	3/36	29
11	2/36	17
12	1/36	8

4. 在非律宾的国际稻谷研究协会(The International Rice Research Institute)开发一批新稻种,将高产、抗病害和抗虫害相结合。技术方面包括杂交各种不同的稻种以获得具有最有利基因组合的新品种;需要详细的遗传学模型。某项科研项目涉及繁殖新品种,以抗一种称为“棕色作物蚱蜢”害虫;共培育了374个稻种,结果如下[15]:

	稻种数
所有作物都有抵抗力	97
混合:一些作物有抵抗力,一些易受虫害	184
所有作物都易受虫害	83

根据IRRI的模型,稻种是相互独立的;每种稻种有25%机会是有抵抗力的;有50%机会是混合的;有25%机会是易受虫害的。事实与这模型相符吗?

5. 两人试图确定一颗骰子是否公平。他们将它掷100次,结果给出如下。一个人要做Z-检验,另一个人要做χ^2-检验。谁对?简短地解释之。

21个 ⚀ 15个 ⚁ 13个 ⚂ 17个 ⚃ 19个 ⚄ 15个 ⚅

掷出点数的平均数≈3.43,SD≈1.76。

6. (a)图2中的直方图代表数据,还是机会?

(b)区间5到5.2上有一块形、这一块形的面积表示什么?(区域包括左端点,但不包括右端点。)

(c)那一个的机会较大些?或者这能用图2决定吗?简短地解释之。

(i)χ^2-统计量在4.8到5.0范围内的机会。

(ii)χ^2-统计量在5.0到5.2范围内的机会。

7. 某调研人员做χ^2-检验,以查看观察频数是否离期望频数太远。若$\chi^2\approx15$,自由度为____的P值将大于自由度为____的P值。供选择的自由度:5,9。(无需计算,只要看图1就行。)

6. 小结

1. 关于数据是根据某特定机会模型产生的这一假设，可以用 χ^2-统计量检验。

2. $$\chi^2 = \frac{(\text{观察频数}-\text{期望频数})^2}{\text{期望频数}}\text{之和}$$

3. 若模型完全确定时（没有参数需要用数据估计）。

$$\text{自由度}=\text{项数}-1$$

4. 观察显著水平 P 可以近似取为 χ^2 曲线之下，χ^2 观察值右面的面积。它给出了由模型产生的观察频数偏离期望频数如同手头那些观察频数偏离期望频数一样远或甚至更远的机会，偏离距离由 χ^2 度量。

5. 有时模型可以看成是真的，问题是要决定数据是否是捏造出来使得观察频数更接近于期望频数。这时，P 值计算为 χ^2 的观察值左面的面积。

6. 如果试验是相互独立的，则 χ^2-统计量可以按加法合并。自由度也只要加起来则可。

7. χ^2-统计量也可用于检验独立性。当数据从一个简单随机样本获得时，而所需要的是有关总体的某一推断。这是合理的，一张 $m\times n$ 的表（没有关于概率的附加约束），有 $(m-1)\times(n-1)$ 个自由度。

29
显著性检验的更准确的考虑

法则的不幸之一(就是)将思想包裹在词藻之中并在此之后一段很长的时间里不再激起进一步的分析

——Oliver Wendell Holmes,JR (美国,1841—1935)[①]

1. 结果显著吗?

在你拒绝原假设之前P必须取到多小?这一点在第26章第4节中报告过,许多统计学家把界限画在5%和1%。若P小于5%,结果是"统计显著的,"且以水平5%拒绝"原假设";若P小于1%,结果是"高度显著的"。然而,这个问题几乎就象你被授权声称"它是冷的"之前,询问它必须达到多冷。温度在70°F是暖的,-20°F确实冷,没有明确的界线。

逻辑上,它与检验相同。在很可能的与不大可能的结果之间没有明确的界线。P值5.1%和4.9%意味着几乎相同的东西。但是,这两个P值可能受到完全不同的待遇,因为许多刊物将只发表那些"统计显著"——5%界线的结果。而一些更有声望的刊物将只发

表“高度显著”——1%界线的结果[②]。这些专断的界线被如此认真地接受，致使许多调研人员仅仅把他们的结果报告成“统计显著的”或“高度显著的。”他们甚至懒得告诉你P的取值，更不必说他们用的是什么检验方法了。

调研人员应该概括数据，说明用什么检验，并报告P—值，不只是将P与5%或1%相比较。

历史注释。 5%和1%界线是从那里来的？为了查清楚，我们必须看一看统计用表的设计方式，t-表是一个好例子（第26章第6节）。它的一部分复制在下面为表1。

表1 一张简化的t-表。

自由度	10%	5%	1%
1	3.08	6.31	31.82
2	1.89	2.92	6.96
3	1.64	2.35	4.54
4	1.53	2.13	3.75
5	1.48	2.02	3.36

在检验中，如何使用这张表？假设调研人员正在做一个自由度为3的t-检验。他们用5%界线，想知道t统计量必须多大才能达到“统计显著”，——一个5%以下的P值，表被设计得使这不费力。他们查看横向3个自由度的一行，竖向5%的一列，在表的主体部分查到2.35；3个自由度曲线之下2.35右面的面积是5%。因此，一旦t大于2.35，则结果是“统计显著的。”换句话说，这张表给出了“统计显著性”的阈值。类似地，它也给出了1%界线的，或横列于表首的其它显著水平的阈值。

R.A.Fisher是首先发表这类表的其中一个，将这类表设计成这种形式似乎是他的主意。在一页上只含有限的篇幅。一旦水平数值被限定，5%显著和1%显著突出为整齐的舍入数，它们很快获得了它们魔力般的生命力。在计算机随处皆是的今天，这类表几

乎被废弃。5%和1%水平亦然。[③]。

习题 A

1. 是或非，并解释之：

 (a)若 P=1.1%，结果是"统计显著的"但不是"高度显著的。"

 (b)若 P=1%的 0.9，结果是"高度显著的"。

2. 是或非，并解释之：

 (a)一个检验的 P 值是原假设为真的机会。

 (b)若一个结果是统计显著的，则它有百分之五的机会应归因于机会，百分之九十五的机会是实际存在的。

这些习题的答案在第 732 页上。

2. 数据窥探

显著性检验的用途是帮助区分实际存在的差与机会变异。因此，人们有时冒然下结论称一个统计显著的结果不可能解释为机会变异，这是不正确的。抽得数的平均数偶而也会高出盒子的平均数 2 个或更多个 SE——仅由于机会。

因此，即使原假设是对的，仍有 5%的机会取到一个由检验称之为"统计显著的"差。这 5%机会可能发生在你的身上：一个不太可能的事件。但不是不可能的。类似地，在原假设下，仅靠侥幸有 1%的机会取到高度显著的差。

一个调研人员做 100 次检验，即使在每次情况中原假设都是对的，他可期望获得五次"统计显著的"结果和一次"高度显著的"结果，——故每次的差只是侥幸(参见 525 页上习题 5 和 6。)正可谓常在河边走，那能不湿鞋。毫无疑问地，没有什么办法去确定差到底是实际存在的还是仅仅归因于机会。

更糟的是，调研人员常常只是在他们看过数据后才决定那些假设要做检验。统计学家称这为数据窥探。调研人员应当说明在统计显著的差出现之前他们做了多少次检验，要想减少被"统计显著的"侥幸结果欺骗的机会，他们应该用一组独立的数据来验证他

们的结论——譬如重复试验[4]。这一忠告很少被听从。

例 1. 整群. 肝癌是一种罕见癌症，常认为是环境缘故引起的。一个有 10 000 居民的镇上，假设某年里有两个或更多病例的机会是小的——大概是 1% 的 1/2。一个整群肝癌病例（在一个小的社区里的若干病例）的出现将促使探究其原因，譬如由于合成的化学药品引起的供水污染[5]。

讨论. 对（假设）100 个这种规模的城镇和某 10 年期间，仅由机会可能产生若干整群。有 100×10=1 000 个镇和年份的组合；而 0.005×1 000=5。如果你不断地检验原假设，迟早你将获得显著的差。

在数据窥探的一个相对较次要的方面已倾注了过多的笔墨：是用单尾的还是双尾的检验。这要害问题用一个虚构的例子是极易看出的。假设要检验一枚硬币是否公平：它按 50% 的概率出现头像吗？抛这枚硬币 100 次，头像出现了 61 次。如果硬币是公平的，头像的期望次数是 50，故 61 与 50 间的差只代表机会变异。为检验这个原假设，需要一个盒子模型。这个模型由从盒子

?? 张 0   ?? 张 1

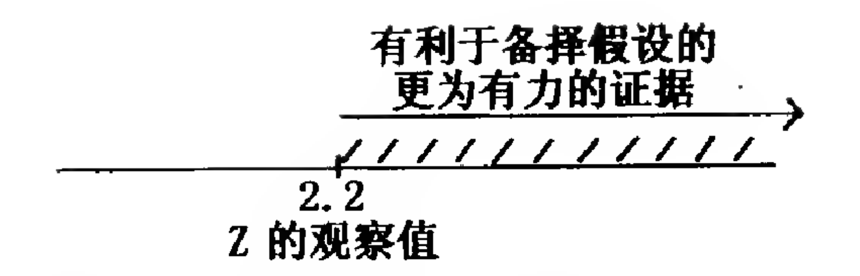
0=背面， 1=头像

中抽取 100 次构成

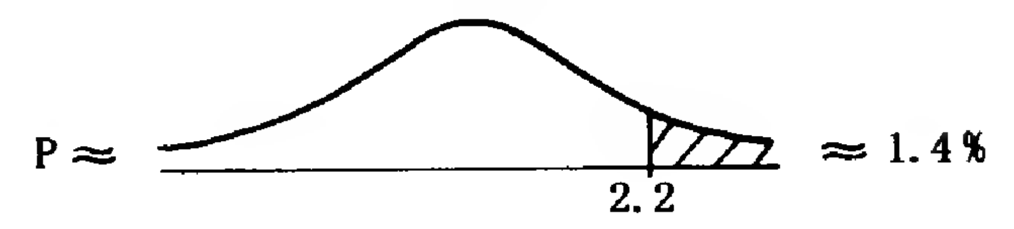
这盒子中 1 占的比率是一个未知数，它代表呈现头像的概率，原假设称盒子中 1 占的比率是 1/2，检验统计量为

$$Z=\frac{\text{观察的}-\text{期望的}}{\text{SE}}=\frac{61-50}{5}=2.2$$

某调研人员可能提出备择假设，这枚硬币偏向头像；换句话说，即盒子中 1 占的比例大于 1/2。在这基础上，大的正 Z 值赞成备择假设，而负的 Z 值不赞成。因此，大于 2.2 的 Z 值比观察值更加赞成备择假设。



故 P 估计为正态曲线之下 2.2 右面的面积：

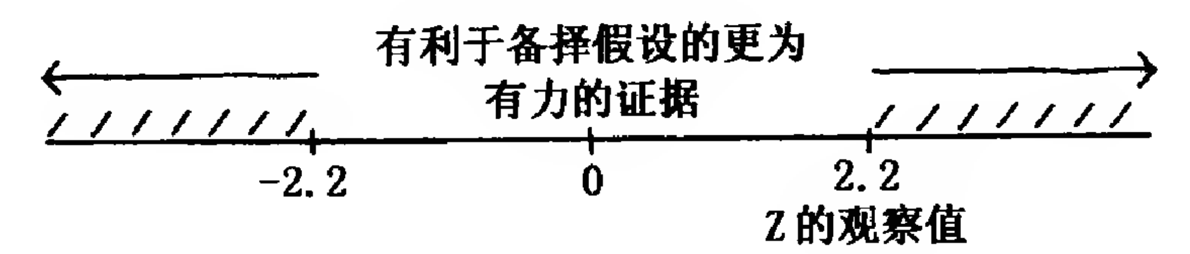


另一调研人员可能提出不同的备择假设：头像的概率在任一方向都不同于 50%。换句话说，盒子中 1 占的比率不同于 1/2，可能比它大或比它小。在这基础上，大的正 Z 值赞成备择假设，并且绝对值大的负值亦然。若头像数在期望值 50 以上 2.2 个 SE，那对原假设不利，若头像数在期望值以下 2.2 个 SE，那也一样不利。比观察的 2.2 更极端的 Z 值是：

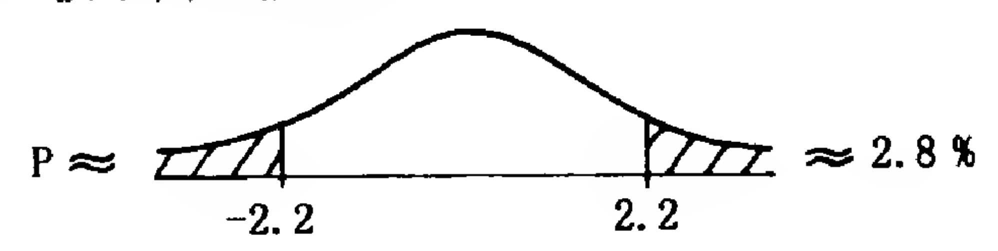
- 2.2 或更大

或

- −2.2 或更小。

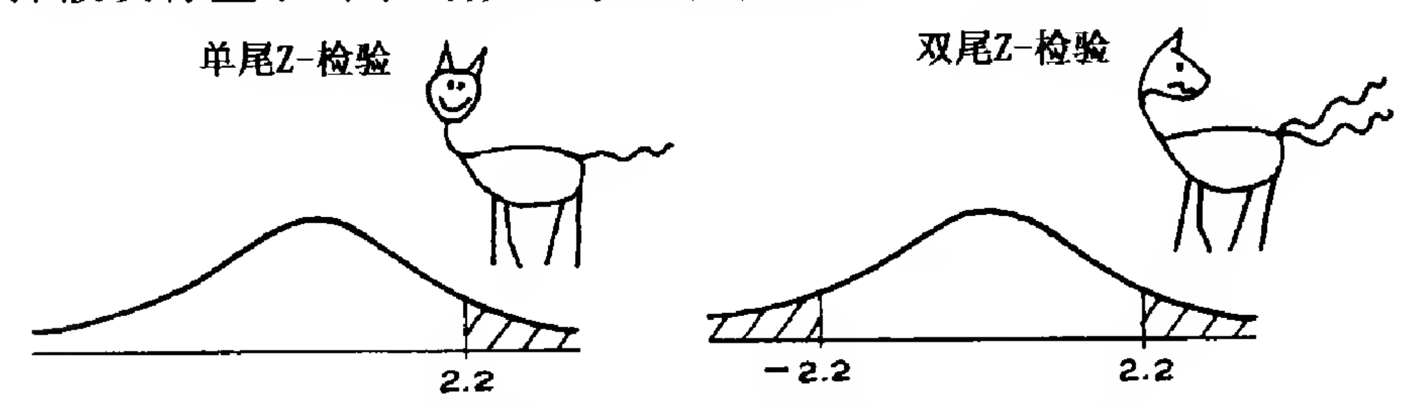


现在 P 被不同地估计为：

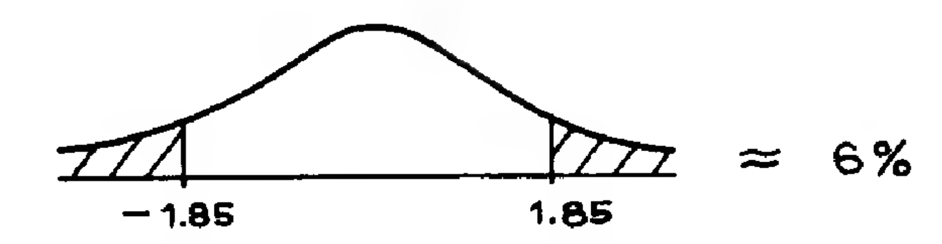


估计 P 的第一种方法是单尾 Z—检验；第二种是双尾的，应该用哪一种？那依赖于备择假设的确切形式。它关系到要看哪些 Z—值比起数据算出的 Z—值更为有力地赞成备择假设。单尾检验适

用于备择假设称盒子的平均数大于某给定值。双尾检验适用于备择假设称盒子的平均数不等于某给定值——大于或小于。

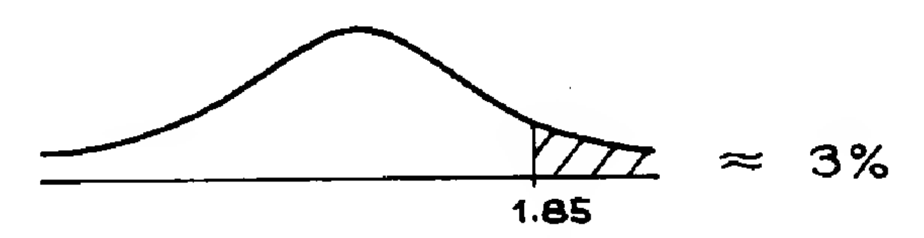


实际上，调研人员是做单尾或双尾检验关系并不大，只要他们讲清楚他们做得是哪一种。譬如，如果他们做的是单尾检验，而你认为它应该是双尾的，则只需将 P—值翻倍则可[⑥]。为了了解为什么对这问题大惊小怪，假设一组调研人员做了一个双尾 Z—检验。他们获得 Z=1.85，故 P≈6%



他们自然想发表他们的结果，但是由于他们维持原状，大多数刊物不会接受这份报告，——结果不是"统计显著的"。

他们能做些什么？可以改进试验方法，收集更多的数据，用更灵敏的分析方法。这样很困难，另一种可能较为简单：做一次单尾检验。



是 5%和 1%这些专断的界线使得双尾和单尾检验之间的区别赫然呈现得如此之大。

例 2. 胆固醇. 二十世纪七十年代，组织了一次随机对照双盲试验以论证一种称为"Cholestyramine"的药在降低血清胆固醇和

防止心脏病突发方面的功效。有 3 806 名实验对象，他们全是处于心脏病突发高险的中年男子；随机选出 1 906 人供处理组，剩下的 1 900 人指派在对照组。对实验对象追踪 7 年，这种药降低了处理组中的胆固醇水平（约 8%）。另外，在处理组中有 155 例心脏病突发，而对照组是 187 例：8.1%对 9.8%，$Z \approx -1.84$，$P \approx 3.2\%$（单尾）。这结果被认为是胆固醇有助于引起心脏病突发的"有力证据。"⑦

讨论. 用双尾检验，$P \approx 6.4\%$，这个差不是显著的。（这篇文章发表在 Journal of American Medical Association（美国医学协会期刊）上，它的编委们对 5%界线是颇为严格要求的。）处理真的能实际上增加了心脏病突发的风险吗？如果如此，单尾检验就是骗人的。这种药似乎还增加了由冠心病之外的病因引起的死亡率：处理组中的 1.9%对比对照组中的 1.4%，虽然这个差可能应归因于机会（$Z = 1.25$，$P = 1.1\%$，单尾）。调研人员在夸大他们的结果，对"统计显著性"的强调鼓励了他们这样做。

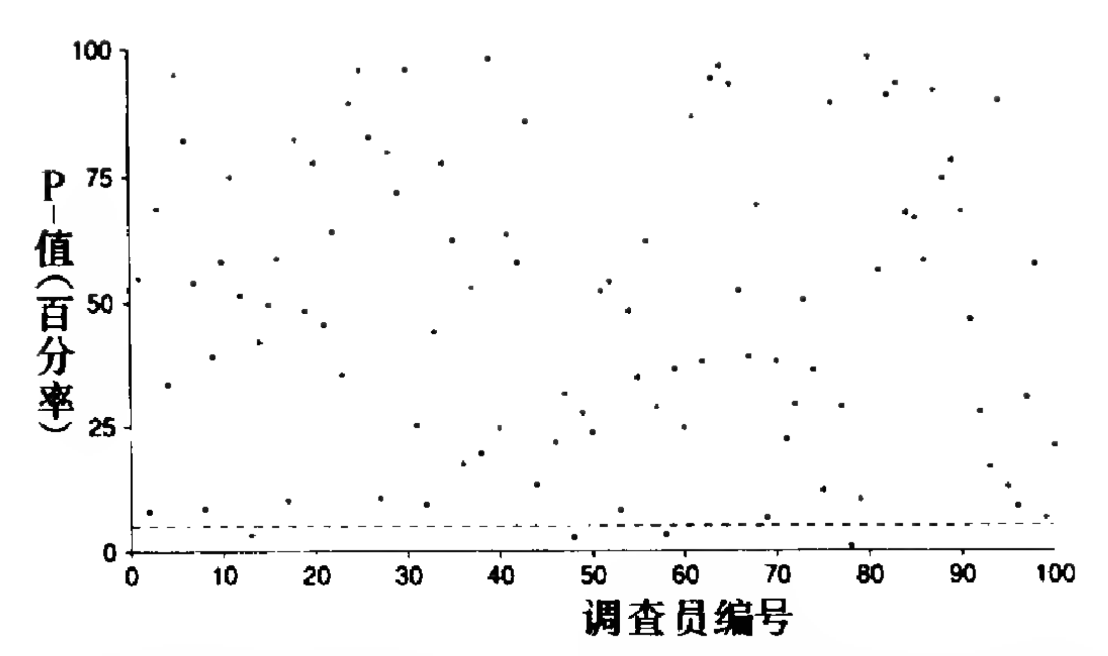
习题 B

1. 一百个调研人员各打算检验一个不同的原假设，他们并不知道，所有的原假设恰好全为真的。调研员 1 获得一个 55%的 P 值，画在下页图中为点(1,55)，调研员 2 获得一个 8%的 P 值，画在(2,8)，余类推。5%界线也在图中标出。
   (a)多少名调研员应该获得统计显著的结果？
   (b)多少名获得了？
   (c)多少名应该获得高度显著的结果？

给你下面信息，供以下的习题使用。假设从盒子 [1][2][3][4] 中随机放回地抽取 25 张票。

- 获得 11 张或以上的编号 1 票的机会约为 3%。
- 获得 17 张或以上的编号 1 或 2 票机会约为 5%。
- 获得抽得数之和为 53 或小于 53 的机会约为 5%。

2. 关于 ESP（超感觉力）的 Ganzfeld"实验中，有两个实验对象，发送者和接受



者，他们被安置在分开的房间里[8]。有一套标准样品图案，整理成25组，每组4张。实验人员挨次试遍这25组。从每组中随机取出一张样品图案，并出示给发送者(但不给接受者)。发送者试图将这张样品图案在他内心中的图象传送给接受者。将组中的四张图案出示给接受者，他按照1=最可能的到4=最不可能的将它们排序。试遍所有25组图案后，实验人员做统计分析，目的是查看接受者是否做得比机会水平好。用了三种检验统计量：

- "击中"次数. 若接受者将序号1指定给实际选出的那张图案则得一次击中。击中次数的取值范围是0到25。(如果击中次数大，那就是ESP的证据。
- "高等级"次数. 若接受者将序号1或2指定给实际选出的那张图案则得高等级。高等级次数的取值范围是0到25。(如果高等级次数大，那就是ESP的证据。)
- 指定给实际选出的25张图案的序号和。这个和数的取值范围是25到100。(若这和数小，那就是ESP的证据。)

假设不存在ESP，也不弄虚做假，并且对样品图案的选择是完全随机的。

(a)击中次数就象取自盒子 | 1 | 0 | 0 | 0 | 的________个抽得数之和。填空并解释。

(b)高等级的次数就象取自盒子________的25个抽得数之和。填空并解释。

(c)构造一个序号和的盒子模型。

3.(续习题2.)假设不存在ESP，也不弄虚做假，并且对样品图案的选取是完全随机的。

(a)一百名调研人员做Ganzfeld实验。若击中次数在11次或以上，则他们

将发表 ESP 的“显著”证据。他们中约有多少人将获得显著证据？

(b)如果将显著证据的定义改为“高等级次数为 17 或更多。”重复(a)的问题

(c)如果将显著证据的定义改为“序号和等于或小于 53。”重复(a)的问题

4. (续习题 2 和 3。)假设不存在 ESP，也不弄虚做假，并且对样品图案的选取是完全随机的。一百名调研人员做 Ganzfeld 实验。他们要在看过数据后才确定统计检验。

- 若击中次数是 11 或更多，他们将检验基于击中次数上。
- 若不，但高等级次数是 17 或更多，他们将检验基于高等级次数上。
- 若不，但序号和小于或等于 53，他们将检验基于序号和上。

这些调研人员中获得 ESP“显著”证据的人数将______5。在空白处填入下面回答之一，并简短地解释。

差不多等于　　略大于　　略小于

5. 审核一些化学药品，以查看它们是否会在实验用小鼠中引起癌症。有一种“生物鉴定”方法可以用 500 只小鼠进行；随机从中选出 250 只，并给它们喂掺有被检测化学药品的食物；另外的 250 只喂以正常的实验室食物。33 个月后，用两样本 Z 检验比较两个组中的癌症率。[9]调研人员查看了约 25 种器官——肝，肺，骨髓等的癌症率。食用一种化学药品，对于肝，$Z\approx2.4$，对肺，$Z\approx-1.8$，对白血病，$Z\approx-2.1$，还有另外 22 个 Z 值取值从 -1.6 到 $+1.5$ 范围内。这些调研人员得出结论称这种化学药品引起肝癌($Z\approx2.4$，$P\approx1\%$，单尾)。简短地评论统计检验的使用。

6. 从盒子 X 中随机抽取 100 次。抽得数的平均数是 5.18，它们的 SD 是 9，原假设称盒子的平均数等于 50，而备择假设称盒子的平均数不等于 50。是单尾或双尾 Z—检验更为适用。

7. 对备择假设为盒子的平均数大于 50，重做习题 6。

8. 某调研人员有取自盒子 A 和取自盒子 B 的独立样本。她的原假设称这两只盒子有相同的平均数。她考虑差

取自 A 的样本的平均数－取自 B 的样本的平均数。

两样本 Z 检验给出 $Z\approx1.79$。

(a)若备择假设称盒子 A 的平均数大于盒子 B 的平均数，该差统计显著吗？

(b)若备择假设称盒子 A 的平均数小于盒子 B 的平均数。该差统计显著

吗？

(c)若备择假设称盒子A的平均数不等于盒子B的平均数，该差统计显著吗？

9.(难题)输血时用污染的血浆将引起感染传染病的危险。(爱滋病就是恰当的例子。)一个内科医生必须在输血的得益与感染风险之间进行权衡，因此精确的数据是重要的。在对已出版的关于输血导致血清肝炎的医学文献的概括研究中，Chalmers 和助手们发现较大的研究有较低的死亡率[10]。这能如何解释？

这些习题的答案在第732页上。

3. 结果重要吗?

如果差是统计显著的，那就很难将它解释为机会变异。但是在这个术语中，“显著”并不意指“重要”，统计显著性与实际显著性是两个不同的概念[11]。在虚构的例子(基于552页上的习题3)的上下文中这点是极易理解的。假设调研人员打算比较大城市和农村6到9岁儿童的WISC词汇量等级得分。他们取一个2 500名大城市儿童的简单随机样本，和另一个独立的2 500名农村儿童的简单随机样本。大城市儿童考试平均得26分，他们的SD是10分；农村儿童平均只得25分，具有相同的SD10分。这一分之差意味着什么？为了找出答案，调研人员做两样本Z—检验。差的SE可估计为0.3，故

$$Z=\frac{26-25}{0.3}\approx 3.3, \qquad P\approx 5/10\ 000。$$

大城市儿童与农村儿童间的差异高度显著，农村儿童在语言能力的开发方面落在后面，调研人员发动一场改革以向农村的学校投注资金。

常识反应必定是：放慢点。Z—检验只告诉我们，样本平均数间的一分之差几乎不大可能解释为机会变异。为了使问题集中，假设样本是总体的一个完全写照，因此全美国所有的大城市儿童(不仅是样本里的那些)的WISC词汇量等级得分平均为26分，而全

美国所有的农村儿童的得分平均数是25分。于是怎么样呢？不再存在为之烦恼的机会变异，故显著性检验不会有帮助。所有的实际情况都在其中，问题是找出差的含意是什么。

要做到这一点，有必要考虑WISC词汇量等级本身。有40个单词是儿童必须掌握的。正确解说得2分，部分正确解说得1分。因此大城市与农村儿童间的一分之差只不过等于部分理解40个单词中的一个。这不是改革的坚实基础。完全相反：调研人员证明了关于WISC词汇量等级，大城市和农村儿童之间几乎没有差别[12]。

当然，样本并没有完全反映总体，因此应该对差的估计附上标准误差。基于两个2 500名儿童的样本，全美国所有大城市与农村儿童间平均得分差可估计为1±0.3分左右。（因为差被如此精确地估计，故Z—统计量为3.3。）

大的样本是好的，因为它使得调研人员可以十分精确地度量差——具有小的SE。但是Z—检验是将差与它的SE相比较；因此，对于大的样本，即使差很小也可能导至引人注目的Z值。Z—检验由于它本身的优点而可能过份敏感。

> 检验的P值依赖于样本容量。对于大的样本，就是小的差也可能是"统计显著的"。这很难用抽取的运气来解释。不必说它是重要的。相反，如果样本太小，一个重要的差可能是统计不显著的。

例3. 如第27章第2节中所述。数学考试的成绩从1973年的55.0分降为1982年的51.8分。这些是基于全国性样本的平均数：$Z \approx -4.2$，$P \approx 1/100\ 000$（单尾的）。下降关系重大吗？

解：P值表明下降难于解释为机会变异。但它没说下降是否关系重大。更详细的数据分析提出高中学生修的每门额外数学课程在它们的平均考分中提高了约6分[13]。在这基础上，3分的下降相当于丢掉半门课程。长远说来，那可能意味着困难。

习题 C

1. 是或非，并解释：

   (a) 一个高度显著的差，必定是非常主要的。

   (b) 大的样本是差的。

2. 某所大型大学打算比较男、女大学生的标准化阅读考试成绩，但是只有能力在一个样本基础上进行。某调查员随机选了 100 名男大学生，并独立选了 100 名女大学生。男的平均考试成绩是 49 分，他们的 SD 是 10 分。女的平均考试成绩是 51 分，SD 同样也是 10 分。平均得分之间的差是实际存在的。还是机会变异？或这问题有意义吗？

3. 重复习题 2，保持平均数和 SD 不变，但将样本容量由 100 增大到 400。

4. 某人解释显著性检验的要点如下[14]。"若原假设被拒绝，差不是不重要的。它比仅由机会所能产生的差大。"简短地评论之。

5. 其它情况都相等，哪一种是原假设的更有力证据：$P=3\%$ 或 $P=97\%$？

6. 文章在学术刊物上发表之前都要经过评审。这个过程公正吗？为了做出解答，某心理学家将一篇叙述研究奖励儿童课堂行为的作用的论文起草成两种描述形式。这两个描述形式是等价的，仅数据不同：一组数据表明奖励有助于促进学习；另一组，奖励无帮助。随机选出一些评审员审阅每一描述形式。他们全体与主张为"行为主义者"的某刊物有关系：奖励对学习应该有作用[15]。

   (a) 某些评审员没有及时返回他们的评审意见。心理学家评论说，"不误期的评审员的推荐意见比迟到的评审意见更接近否定的边缘，$P\approx 4\%$，单尾的。"这结论是统计检验的结果吗？

   (b) 正如结果所示，论文的两种描述形式都包含了研究描述中一处较次要的前后自相矛盾的内容。"接到肯定描述形式的人中仅 25% 发现这一错误。接到否定描述形式的人中有 71.5% 发现这一错误。根据两样本 Z—检验，这个差必须看成是本质的，$P\approx 20\%$，单尾的。"这结论是统计检验的结果吗？

7. 某经济学家通过在 San Francico（旧金山）和 Los Angles（洛杉机）的 California Divison Of Highway（加州公路部门）进行了一项关于快速干道选择的研究项目[16]。他提出一个机会模型以便获取各种不同变量对决策的影响。"客观的政治和公众变量"包括其它的州机构、学校部门、商行、大量资

产拥有者、以及资产拥有者协会的意见。为了查清这些变量对快速干道决策有多大影响，该经济学家做了显著性检验，原假设称客观的政治和公众变量在决策中不起作用。观察的显著水平约为3%。由于结果是统计显著的，但是不是高度显著的，因此该经济学家下结论说“这些因子影响快速干道的决策，但是它们的影响力是相对弱的。”这结论是统计检验的结果吗？

8. 某经济学家估计精炼油产品的价格顺应性为－6。(－6的顺应性意指价格上升1%导至销售下降6%。)标准误差是2.5，他检验原假设：顺应性为0，并获得 $Z=-6/2.5=-2.4$ 和 $P\approx1\%$(单尾)。结论：他“99%信任估计值。”[17]简短地评论检验的使用和置信水平。

这些习题的答案在第733页上。

4. 差证明论点吗？

通常调研人员收集数据以证明论点。如果结果可由机会解释，数据不能证明任何东西。因此，调研人员做显著性检验以证明差是实际存在的。但那是检验所能做的一切。即使差是实际存在的，它仍然不能证明论点。显著性检验可以排除机会作为差的似乎可能的解释。但它们不能回答这样的问题，“什么东西引起这个差？”

以ESP实验为例，在实验中掷一颗骰子，实验对象试图使它出现六点[18]。将这重复720次，这些试验的143次中骰子出现六点。如果骰子是公平的，而且实验对象的努力无作用，骰子有1/6的机会出现六点。因此，在720次试验中，六点的期望次数是120。有剩余额143－120＝20个六点。

这个差是实际存在的，还是机会变异？这正是显著性检验该解决的。原假设可根据这样的盒子模型而提出：六点的次数就好象从盒子 0 0 0 0 0 1 中抽取的720个数之和。抽得数之和的SE是 $\sqrt{720}\times\sqrt{1/6\times5/6}=10$，故 $Z=(143-120)/10=2.3$，$P\approx1\%$。这个差看来是实际存在的。

现在，另一个问题：这个差证明ESP存在吗？毕竟，这是关键。实际结果是，在另外一部分试验中，实验对象试图使骰子出现幺

点，而获得过多的六点。事实上，无论实验对象试着要几点，都有过多的六点。这不证明 ESP 存在。而证明了骰子是有偏的。过多六点的原因不是 ESP，而是骰子的偏性。

显著性检验不能告诉你差的起因；它只能告诉你差在那里。这把我们带回到第一部分中所讨论的设计问题。对于一个良好设计的实验，实际存在的差证实了调研人员的论点。对设计上有缺点的研究，较难得出结论。

显著性检验不检验研究的设计

在 ESP 实验中，为什么 Z—检验把我们引入岐途？不是这样的。检验被问道是否存在太多的六点以致不能用机会来解释。而它正确地回答，是多了点。可是检验被告知“机会”指什么意思：掷一颗公平的骰子。这个假设被用来计算 Z 的公式中的期望值和 SE。显著性检验必须报被告知使用何种机会。如果一个调研人员误解了机会，如 ESP 实验中那样，不要责怪检验。下一节将有更多关于这方面的讨论。

例 4. 第 26 章第 5 节讨论了 Tart 的 ESP 实验。一台称为“Aquarius（宝瓶座）”的机器随机从四个目标挑出一个，实验对象试着猜是那一个。在 7 500 次试猜中实验对象猜中 2 006 次，比较之下机会水平有 $1/4\times7\ 000=1\ 875$ 次，差是 $2\ 006-1\ 875=131$，$Z\approx3.5$，$P\approx2/10\ 000$（单尾）。这证明了什么？

讨论. 这个差难于解释成机会变异。那就是 Z—检验所给出的。但它是 ESP 吗？为排除其它的解释，我们必须考察研究的设计。结果发表后，统计学家们检查了 Tart 使用的随机数发生器。这些随机数发生器有一个缺陷：它们很少在一排里两次中挑出同一目标。在实验中，实验对象做出每一猜测后，Aquarius 机器都照亮它选出的目标。那些注意到图案或由于其它原因每次挑选新目标的实验对象，可能由于随机数发生器的缺陷而改善他们的机会。兴许它终究不是 ESP。

Tart 先是否认这种替代解释。最后，他用更好的随机数发生

器重复该实验。在其它方面，他也严格设计。在重复试验中，实验对象以大约于机会水平猜中：不存在 ESP。这两次实验的实验对象都是 California Davis 大学的学生。Tart 对重复实验遭到失败的主要解释：两次试验之间学生表态的“某种戏剧性变化。”[19]

> 在过去的一两年里，学生们比起他们在第一次实验的时候变得更认真、更富竞争性和更着重成就。这种“更为拘谨的”态度与以强烈的兴趣和动力去探索与推动一种诸如 ESP 的“无用”的才能较少一致。的确，我们注意到在目前的试验中我们相当一部分参与者似乎没有真正“投入”实验，而只是急于“草草了事。”

习题 D

1. 524 页上的习题 4 讨论了一项实验，该实验将弹性工作时间介绍到一个工厂，在一个样本的基础上，旷工由平均 6.3 天不工作降为 5.5 天。一次检验称这个差是实际存在的。得出结论说弹性工作时间造成这个差合理吗？若不，一些其它的可能解释是什么？如何才能改进这项实验的设计？
2. 第一章讨论了脊髓灰质炎疫苗现场实验，其中疫苗组比对照组有少得多的小儿麻痹症病例。一项显著性检验指出差是实际存在的（562 页上的习题 2）。得出疫苗防止了小孩患小儿麻痹症的结论合理吗？若不，一些其它的可能解释是什么？
3. 糖精作为一种人造低热量甜化物被用于减肥软饮料。但是担心它可能致癌。调研人员在小鼠身上进行了生物鉴定。（生物鉴定曾在 599 页习题 5 中讨论。）在处理组，动物获得它们每日食物的 2% 中摄入糖精。与对照组相比，有较高比率的膀胱癌。差是高度显著的。调研人员得出结论称糖精对人很可能引起癌症。这是解释 P—值的一个好的方式吗？
4. 某公司有 7 名男雇员和 16 名女雇员。可是男人比妇女挣钱多，公司被指控在工资制定中存在性别歧视。原告的专家推理如下：

   > 有 7×16=112 对雇员，每对中有一个是男性，第二个是女性。在这些对中有 68 对男人挣得较多。如果不存在性别歧视，男人应该只有 50—50 机会挣得较多。这就象抛一枚硬币，在抛一枚硬币 112 次中，头像的期望次数是 56，以及一个约 5.3 的 SE。因此

$$Z=\frac{\text{观察的}-\text{期望的}}{\text{SE}}=\frac{68-56}{5.3}\approx 2.3$$

P≈1%。如果我在任何时候见过它。那就是性别歧视的证明。

你同意吗？回答是或否，并予解释。

这些习题的答案在第733—734页上。

5. 模型的作用

简短地回顾，显著性检验回答这样的问题，“差归因于机会吗？”但是，除非给“机会”这个词予精确定义，否则检验不可能发挥它的作用。这就是盒子模型该做的份儿。[20]

要使显著性检验有意义，需要一个盒子模型。

这想法可能有点令人惊奇，因为检验的算术运算中没有用到盒子模型。代之于，检验似乎直接从数据产生机会，然而，那是一个错觉。就是盒子模型定义了机会。期望值和标准误差的公式作出了一个心照不宣的假设：即数据就象取自盒子的抽得数。统计用表——正态、t和χ^2也如此，如果这个假设是错误的，公式和表不再适用，并且还可能给出愚蠢的结果。本节将讨论一些例子。

例5. 人口普查数据表明，1970年美国有2亿零3百万人，其中9.8%是65岁或更老。1980年有2亿2千7百万人，其中11.3%是65岁或更老。[21]百分数之差统计显著吗？

讨论. 两样本Z—检验的算术运算相当容易进行（特别是用计算器），但是结果近乎无意义。我们有全部人口的人口数据，不需要担心抽样变异性，人口普查数据带有许多小的误差，但是这些并不象从盒子中的抽取。人口的老化是实际的，重要的。但是，统计显著性概念不适用。

如果显著性检验是基于整个总体的数据，请小心一点。

例6. California大学，Berkeley分校研究生部的记录可以用来比较男女的录取率。有一年某研究生专业的实际情况如下：825名男性提出申请，61.7%被录取、108名女性提出申请，82.4%被

录取[22]。男女录取率之间的差统计显著吗?

讨论. 再一次,没有什么可以阻止你做两样本 Z—检验。但是,要想使结果有意义,将需要一个盒子模型,而在邻近似乎并不存在一个盒子模型。机会可能进入吸收申请者,或者进入确定录取那个申请者。然而几乎不可能去识别有可能性的申请者的全体,即使你能做到,实际申请者也不是以某种概率方法从这个集合中抽取,各个系也不是用从一顶帽子里抽取名字的方法录取候选人(尽管这可能并非是一种坏主意)。统计显著性概念不适用。

统计学家把依概率方法抽取的样本和方便的样本已分开(第22章第5节)。方便的样本由近便的任何人组成——一年级心理学班级的学生,偶然遇到的前100人,或某特定年份申请某系的全体申请者。对于方便的样本,机会这个概念变得非常难于捉摸:"差归因于机会"这句话难于解释,P—值也如此。例6就是基于一个方便样本的。

如果显著性检验基于一个方便的样本。请小心一点。

例7. 一贯到底方案(Project Follow Through)是国会制定的,目的是为 Headstart 学龄前规划中的少数民族儿童,当他们进入正规学校时做出的学术上的收获提供资助。授予七名赞助人合同以便根据不同的教育哲学管理项目课堂,并以其它一些课堂为对照。Stanford 研究院(SRI)被雇用为健康,教育和福利部评估该项目[23]。一个重要的问题是项目课堂是否真正地不同于对照课堂,当然,SRI 还研究了有关儿童规划的作用。

为了查看是否存在真正的差,SRI 设计了一种补充得分用来对项目课堂与对照课堂做比较,这种得分包括对课堂进行观察并确定,比如,儿童花费在玩、独自工作、向教师提问问题方面的时间,等等。发起者之一的 Far West 研究室的结果给出在表2中。

表 2　Far West 研究室的 20 个课堂的 SRI 补充得分。得分在 0 到 100 之间。

地点	课堂得分			
Berkeley	73	79	76	72
Duluth	76	84	81	80
Lebanon	82	76	84	81
Salt Lake City	81	86	76	80
Tacoma	78	72	78	71

这 20 个得分的平均数约为 78；它们的 SD 约为 4.2。对照课堂的平均得分约为 60；故差是 18 分。就 SRI 的补充得分而论，Far West 的课堂非常不同于对照课堂。到这里为此，非常好。但是 SRI 并不不满足，他们想做 Z-检验，

目的是检验 Follow Through 的平均补充得分是否显著地大于非—Follow Through 的平均数

计算如下[24]。得分之和的 SE 估计为$\sqrt{20}\times 4.2\approx 19$；故它们的平均数的 SE 为 $19/20\approx 1$，且

$$2=(78-60)/1\approx 18$$

于是，



推断为：

Far West 课堂的综合平均数显著地不同于非—Follow Through 课堂数平均数 60。

讨论. 算术运算全都妥当，并且一开始过程似乎也是合理的。但是，存在一个实际问题，因为 SRI 没有一个数据的机会模型。并且他们还难于创立可能合理的一个。SRI 可能认为 20 个处理组课堂是来自所有课堂总体中的一个样本。但是，他们没有按照简单随机抽样选取他们的 20 个课堂，甚至也没有按照某种较复杂的概率

方法选取。实际上，在报告中没有描述选取课堂的明确程序。这是一个纯粹的和简单的、方便的样本。

SRI 可能想到了测量误差。是否存在 Far West 的某种"精确值，"它可以或不可以与对照的那一个不同？如果存在，这是一个单一的数吗？或者它依赖于地点？依赖于课堂？教师？学生？年份？或者这些是误差盒子的一部分吗？如果真如此，误差盒子随课堂的不同，或者随地点的不同而不同吗？误差不是要相依了吗？

报告共 500 页，没有一页触及这些问题。这把在某种外界标准下，可以用显著性检验比较任何样本的平均数，不管它是从那里来的，当成是不言而喻的。证明项目课堂不同于对照的整个讨论都依赖于这些检验，而检验却无可依赖。SRI 没有一个容量为 20 的简单随机样本，也没有关于同一个量的 20 个重复度量。它有 20 个数。这些数含有机会成分，但是对产生它们的机制几乎一无所知，在这些条件下，显著性检验是一种智力绝望的行为。

我们往 SRI 与调研人员讨论这些问题，他们坚持认为，在设计研究时，他们吸取了非常好的统计方面的忠告，并且只不过在做任何其它人做过的事。我们坚持自己的论点。讨论持续了若干小时。最后一位资深的调研员说：

> 看，我们设计这项研究的时候，我们的顾问之一解释，有一天，某人会自蓝天到来并称我们的统计学没有任何意义。所以你瞧，每件事都非常仔细考虑过。

习题 E

1. 某一学期，California 大学 Berkelay 分校有 600 名学生参加了统计学 2 的期终考试。平均得分是 65 分，SD 是 20 分。在紧接着的学年初，指派给这门课的 25 名助教参加了完全相同的考试。助教们的平均得分 72 分，他们的 SD 也是 20 分[25]。助教们显著地比学生们好吗？若合适的话，做两样本 Z—检验。若这是不合适的，解释为什么不合适？
2. 古时代所知道的 5 颗行星可分为二个群体：内行星（水星、金星），它们比地

球更靠近太阳；和外行星（火星、木星、土星），它们更远离太阳。这些行星的密度给出如下（地球的密度取为1）。

水星	金星	火星	木星	土星
0.68	0.94	0.71	0.24	0.12

两颗内行星平均密度是0.81，而三颗外行星的平均密度是0.36，这个差是统计显著的吗？[26]或这问题有意义吗？

3. 两名研究人员研究Pennsylvania州Dauphin县的婴儿死亡率与环境条件之间的关系。作为研究的一部分，研究人员对某一长达6个月的期间在Dauphin县出生的每一个婴儿记录了婴儿的出生季节，以及婴儿是否在一周岁前死亡。[27]如果合适的话，检验查看婴儿死亡率是否与出生季节相关。若检验是不合适的，解释为什么不合适。

	出生季节	
	七、八、九月份	十、十一、十二月份
周岁前死亡	35	7
生存壹年	958	990

4. 在WISC区组设计测验中，发给实验对象彩色木块料，并要求把它们装配成图中所示的各种不同图案。作为第二轮健康检查调查的一部分，对一个按照概率方法抽取的6到9岁儿童的全国性样本进行了这项测验。这个样本基本上是现场人口调查中所采用的那种多阶整群样本（第22章）。在样本中有1 652名儿童其家庭年收入在5 000美元到7 000美元；这些儿童测验平均得14分，它们的SD是8分。在样本中有813名儿童其家庭年收入在10 000美元到15 000美元，这些儿童测验平均得17分，他们的SD是12分[28]。某人问平均数间的差是否可以解释为机会变异。

(a)这问题有意义吗？

(b)它能在给定的信息基础上回答吗？

简短地解释。

5. 政治分析员认为州很有关系：不同的州有不同的政治文化，这些政治文化形成了投票人的看法。[29]经过对某些人口统计变量加以控制后，调研人员估计居住的州对参加党派（共和党或民主党）的影响。数据基础由CBC/New York Times于1976—1982年期间在美国抽样调查的55 145人组

成。原假设——州之间无差异——被拒绝(Z≈17,P≈0 是经过调节供州间交叉多重比较用的)。简短地评论统计检验的使用。

6. 某调研人员询问在 Mc Carthy(麦卡锡)年代对左翼观点的政治压制应归因于“公众意见或名流意见。”[30]他检查了由州立法机构通过压制法令,与有关意见调查——

> …一个公众和政治名流的样本…选出的名流不是州名流随机样本的…相反,名流样本只代表他们自己…公众意见(的影响)是 -0.06;对于名流意见它是 -0.35(超出 0.01 显著的)。因此,政治压制发生在有相对偏执的名流的州里。在名流们的偏执之外,公众选择似乎关系不大。

简短地评论统计检验的使用。

这些习题的答案在第 734 页上。

6. 结 论

> 当事人去接受盘问的时候,律师通常给予如下忠告:
>
> 留心听问题,并回答问题。不要回答他们本应问但未问的问题,或回答你想让他们问的问题。只回答他们真正问的问题。

显著性检验遵循完全不同的策略,无论你问什么,它们回答一个且仅一个问题:

> 只基于机会变异,解释数据与根据原假设所能期望的这两者之间的差有多容易?

机会变异由盒子模型定义,这个模型由调研人员指定(明确地或含蓄地)。检验并不查核这个模型是恰当的或者是似乎可能的。检验将不测量一个差的大小。或它的重要性。它也不鉴别差的起因。因此,检验只能回答一个非常特定的问题。常常那是一个要问及的错误问题。因此这问题不该由检验来解答,而该由估计来解答。这包括构造数据的机会模型,定义你想根据模型来估计的参数,并由数据估计它,同时给估计赋予标准误差。

现今,显著性检验相当普及。一个理由是它是给人予深刻印象

且又发展很好的数学理论的一部分。另一个理由是对许多调研人员不想费心设计机会模型。检验的措词使得人们易于回避模型，而大谈“统计显著”的结果。这听起来那么令人印象深刻，并且在幕后有那么多的数学方法呐喊助威，使得检验似乎确实是科学的——即使在它们毫无意义 的时候。ST. Exupéry 非常理解这类问题：

当难于理解的事物太不可抗拒的时候，人们不敢不服从。

——The Little Prince[31]

7. 复习题

复习题可能包含前面几章的内容。

1. 下列问题中的那一个用显著性检验处理？

   (i)差归因于机会吗？

   (ii)差重要吗？

   (iii)差证明了什么？

   (iv)实验是适当地设计的吗？

   简短地解释。

2. 高度显著的结果仍能归因于机会吗？简短地解释。

3. 两名调研人员在检验有关盒子 X 相同的原假设：即它的平均数等于 50。他们商定了备择假设，即平均数不等于 50。他们也都赞成用双尾的 Z—检验。第一个调研员随机放回地从盒子中取 100 张票。第二个随机取 900 张票子，也是有放回的。两个调研员获得相同的 SD 10。正确还是错误：得到远离 50 的平均数的调研员将获得更小的 P 值。简短地解释之[32]。

4. 在 Title Ⅶ 诉讼中。法庭认为当商行的雇员中黑人的百分数低于周围地区黑人的百分数，只要由 Z-检验得出差异是“统计显著”的，则存在着一个对商号不利的种族岐视案件。假如在某城，10%的人是黑人。又假定该城市每一个商行雇用职工是通过这样一种过程，就考虑到种族而言，等价于简单随机抽样。是否这些商行中的任何一个有可能因为 Z—检验而被发现犯有种

族岐视吗？简短地解释。

5. 内行星（水星、金星）是那些比地球更靠近太阳的行星、外行星是离得更远些的。这些行星的质量给出如下，取地球的质量为1。

水星	金星	火星	木星	土星	天王星	海王星	冥王星
0.05	0.81	0.11	318	95	15	17	0.8

内行星的质量平均为0.43，而外行星的质量平均为74。这个差是统计显著的吗？[33]或这个问题有意义吗？简短地解释。

6. 调研人员利用选举数据，进行了一项有关影响投票行为的各种因素的研究。他们估计在某次选举中，通货膨胀问题使共和党的选票增加了7个百分数点。但是，这个估计的标准误差约为5个百分数点。因此，该增量不是统计显著的。调研人员得出结论说，"事实上，与广泛持有的观点相反，通货膨胀对投票行为无影响。"[34]这个结论是根据统计检验得到的吗？回答是或否，并简短地解释。

7. 根据人口普查数据，1950年美国的人口总数是一亿五千三百三十万人，他们中的13.4%居住在西部。1980年，人口是二亿二千六百五十万人，他们中的19.1%居住在西部。[35]百分数间的差实际上显著吗？统计显著吗？或这个问题有意义吗？简短地解释。

8. 根据现场人口调查数据，在美国，在1975年，年龄在16岁或以上的妇女有42.0%受雇；到了1985年，百分数增长到50.4%[36]。百分数间的差统计显著吗？

   (a)问题有意义吗？

   (b)你能基于所给信息回答这个问题吗？

   (c)如果你假定现场人口调查是基于每年50 000名年龄在16岁或以上妇女的独立简单随机样本，你能回答它吗？

9. 1970年，48%的大学新生认为"已婚妇女的活动最好限于家和家庭。"到了1985年，百分数降为22%[37]。这些百分数是基于全国性多阶整群样本得到的。

(a)这个差重要吗？或这个问题有意义吗？

(b)问该差是否统计显著这个问题有意义吗？你能在已给定信息的基础上回答吗？

(c)重复(b)，假设这些百分数是基于每年抽取 1 000 名新生的独立简单随机样本得到的。

10. M. S. Kanarek 和助手们在 San Francisco(旧金山)湾周围的人口普查局的 722 个调查片研究癌症率与饮用水中石棉标准之间的关系。[38]对年龄和各种人口统计变量，但不包括抽烟，进行调整后，他们发现男性白人中的肺癌率与饮用水中棉纤维的浓度之间有“强相关性”：P<1/1 000。

    石棉浓度乘以因子 100 时，平均来说相应肺癌水平约按因子 1.05 增加。(若 B 片水中石棉纤维浓度是 A 片的 100 倍，又 A 片里男性白人的肺癌率是每年 1000 人之一，则在 B 片中肺癌率预计为每年每 1000 中之 1.05。)

调研人员检验了超过 200 种关系——不同类型的癌症，不同的人口统计群组，可能混淆变量的不同的的调节方法；男性白人中肺癌 的 P 值是他们所获得的最最小的一个。饮用水中的石棉引起癌症吗？影响是强的吗？简短地讨论。

11. R. E. Just 和 W. S. Chern 声称 California(加州)的听装番茄买主们运用市场力确定价格。作为证据，调研人员估计了两个时期(引入机械收割机之前和之后)番茄需求的价格弹性。(比如，一个－5 的价格弹性意思是价格上升 1%导致需求下降 5%。)他们赋予这些估计标准误差。

在一个竞争的市场中，收割机不应该造成需求弹性的不同；它只影响供应。但是，两个估计的弹性之间的差——收割机之前和收割机之后——是统计显著的(Z≈1.56，P≈5.9%，单尾)。调研人员在确定最终看法之前，试了估计价格弹性的若干种方法[39]。简短地评论关于统计检验的使用。

12. Belmont 和 Marolla 进行一项有关出生次序，家庭大小和智力

间关系的研究[40]。实验对象由1963年和1966年之间年满19岁的所有荷兰男人组成。这些男人按法律规定接受荷兰军队入伍检查，其中包括Raven的智力测验。结果表明，对每种家庭大小，测出智力随出生次序下降：老大优于老二，老二优于老三，等等。对任一特定的出生次序，智力随家庭人口增加而下降：比如，二个小孩家庭里的老大优于三个小孩家庭里的老大。这些结果即使在对父母的社会等级控制后仍为真。

譬如，以两子女家庭为例：

- 老大测验平均得2.575分；
- 老二测验平均是2.678分。

(Raven测验得分范围是1到6，1分最好，6分最差。)这个差是小的，但是，如果这个差是实际存在的，它就具有遗传学理论的有趣含义，为了证实差是实际存在的，Belmont和Marolla做了一个两样本Z—检验。老大和老二的测验得分的SD都约为1分，且各有30 000人，故

$$和的\ SE \approx \sqrt{30\ 000} \times 1\ 分 \approx 173\ 分$$

$$平均数的\ SE \approx 173/30\ 000 \approx 0.006\ 分$$

$$差的\ SE \approx \sqrt{(0.006)^2 + (0.006)^2} \approx 0.008\ 分$$

因此，$Z \approx (2.575 - 2.678)/0.008 \approx -13$，P惊人地小。Belmenf和Marolla得出结论说：

> 因此观察差是高度显著的……由于大量的事实，对每个平均数可赋予高水平统计置信度。

他们的统计程序合适吗？回答是或否，并解释之。

8. 小 结

1. 若P小于5%，结果是“统计显著的”；若P小于1%，则是“高度显著的”。不过，这些界线有点随意性，因为在很可能的与不大可能的结果之间没有明确的界线。

2. 调研人员应该概括数据，说明用的是什么检验，并报告P—

值，代替只将P与5%和1%相比较。

3. 即使结果是统计显著的，它仍有可能归因于机会。因此数据窥视使P-值难于解释。

4. Z-检验可以是单尾的或双尾的，取决于备择假设的形式。

5. 检验的P-值依赖于样本容量。对于大样本，即使是小的差也可能是“统计显著的”，即难于解释为机会变异。但这不能说明差异是重要的。相反，一个重要的差，如果样本太小的话，也可能是统计不显著的。

6. 为了确定样本中观察到的差是否重要，假装它适用于整个总体，而查看它在实际上的含义是什么：这就是“真实显著性检验。”

7. 显著性检验研究差是实际存在的还是归因于机会变异这种问题。它不说差有多重要，也不说什么原因引起它。检验也不核查研究的设计。

8. 通常，当数据是整个总体的时，显著性检验是无意义的，因为不存在要求的机会变异。

9. 为检验差是否归因于机会，你必须定义机会，那就是盒子模型该做的，又除非存在数据的一个机会模型，否则显著性检验没什么大意思。特别地，如果检验是用于方便的样本时，P—值可能不表示任何意思。

注释

第一部分 实验设计

第1章 对照实验

1. 比较方法首先可能用于19世纪初，用以证明放血对于肺炎并不是那么有效的治疗方法。见 Pierre Louis，Recherches sur les Effets de la Saignée dans quelques Maladies inflammatoires；et sur l'Action de l'Emétique et des V'esicatoires dans la Pneumonie（Paris：J. B. Baillière，1835；英文版，1836；The Classics of Medicine Library 再版，1986）。有关讨论，见 R. H. Shryock，The Development of Modern Medicine（University of Pennsylvania Press，1936，P. 163）。也应该注意到 Lind 用于维生素C治疗坏血病的试验：见 K. J. Carpenter，The History of Scurvy and Vitamin C（Cambridge University Press，1986）。
2. Thomas Francis，Jr. et al. "An evaluation of the 1954 poliomyelitis vaccine trials—summary report"American Journal of Publie Health vol. 45(1955) pp. 1—63。也可见 P. Meier 的文章。"The biggest publie health experiment ever：the 1954 field trial of the Salk poliomyelitis vaccine ，"在 J. M. Tanur 等的 Statistics："A Guide to the Unknown，3rd ed.（Wadsworth，1989）。
3. "控制你能办到的而随机化其余的。"是由统计学家经常给予的忠告。相配或限制的措施将使方差缩减而付出的代价是使分析复杂化。也见第19章的注释11，和第27章的注释14。
4. H. K. Beecher. Measurement of Subjective Responses（Oxford Unirersity Press. 1959，pp. 66—67）也见 Berton Roueché，The Medical Detectives ，vol . I（New York：Washington Square Press，1984，chapter 9）。
5. N. D. Grace，H. Muench，and T. C. Chalmers，"The present status of shunts for portal hypertension in cirrhosis，"Journal of Gastroenterology vol. 50（1966）pp. 646—91。我们在 J. P. Gilbert，R. J. Light，and F. Mosteller，"Assessing social innovations：an emupirical guide for policy，" Benefit Cost and policy Analysis Annual(1974)中发现了这个例子。
6. T. C. Chalmers，"The impact of controlled trials on the practice of

medicine"Mount Sinai Journal of Medicine vol. XLI(1974)pp. 753—59.

H. Sacks, T. C. Chalmers, and H. Smith, "Randomized versus historieal eontrols for clinieal trials," American Journal of Medicine vol. 72(1982) pp. 233—40. 他们的"随机化对照试验"的定义是不严格的。

第2章 观察研究

1. **参考文献**:

J. Berkson, "The statistical study of association between smoking and lung cancer," Mayo Clinic Proceedings vol. 30(1955)pp. 319—48,

R. A. Fisher, Smoking: The Cancer Controversy(Oliver & Boyd, 1959)。

J. Cornfield, W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder, "Smoking and lung cancer: recent evidence and a discussion of some questions," Journal of the National Cancer Institute vol. 22(1959)pp, 173—203。

U. S. Publie Health Service, Smoking and Health: Report of the Advisory Committee to the Surgeon General(Washington, D. C, 1964)。

International Agency for Research on Cancer, Tobacco Smoking, Monograph 38(Lyon, France: 1986)。

2. The Coronary Drug Project Research Group, "Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project," New England Journal of Medicine vol. 303(1980)pp. 1038—41。其他的药是从1966年三月至1969年十月注册的雌激素(以两次的剂量),右旋甲状腺素,和烟碱酸。由于有害的副作用缘故,雌激素和右旋甲状腺素停止使用。安妥明和烟酸降低了胆固醇水平但不减少死亡率。见"Clofibrate and niacin in coronary heart disease," Journal of the American Medical Association vol. 231(1975)pp. 360—81。但同时见Lancet(三月4日,1989)pp. 473—74,在后来的诊断后与病人保持联系中它表明了烟酸的积极作用。

一个有趣的侧面消息:在基线测量了大约40个危险因素。作为一个组来说,当研究开始时不坚持者似乎处于更差的状态。但是通过使用这些危险因素的回归去彻底地调整坚持者与不坚持者之间的差异是不可能的:对于草案的坚持并不很好地与测量的协变量相关联。这就表明在观察研究中使用回归模型控制混杂因素时要小心。

3. 引文来自于 D. A. Roe，A Plague of Corn (Cornell University Prees，1973)。另外一本重印了许多带有评注的原始论文的出色的参考文献是 K. J. Carpenter，Pellagra(Academic Press，1981)。也可参看 M. Terris，ed，Goldberger on Pellagra(Louisiana State University Prees，1964)。

玉米从美洲被带往欧洲。印第安人在烧玉米之前先用碱处理一下，这样释放出烟酸；糙皮病对他们不成问题。在美国当磨坊主开始从玉米中提取胚芽时糙皮病似乎开始流行；胚芽包括许多可用的烟酸或色氨酸。(色氨酸是一种氨基酸，它在体内可转化为烟酸。)在美国，糙皮病死亡大约在 1930 年达到高峰，自那以后相当平稳地下降，大概因为在经济条件和饮食方面的普遍改善的缘故；在面粉中加料来得太晚以致于没有太大效果。糙皮病在非洲和印度部分地区仍然是地方性疾病。八十年代后期在南非的研究提出了一个可能的毒枝菌素的病因学作用，因此对于它来说传染理论也许很少有真实性。我们感谢 K. J. Carpeter(University of California，Berkeley)在此例中给我们的帮助。

4. **参考文献：**

E. L. Wynder，J. Cornfield，P. D. Schroff，and K. R. Doraiswami，"A Study of environmental factors in carcinoma of the cervix，"American Journal of Obstetrics and Gynecology vol. 68(1954)pp. 1016—52。

更新的依据概述在下述文献：

J. Cairns，Cancer：Science and Society(W. H. Freeman & Co，1978，p. 59)

R. Peto，ed. Yiral Etiology of Cervical Cancer Banbury Reporf no. 21(Cold Spring Harbor，1986)。

S. H. Swan and W. L. Brown。"Oral contraceptive use，sexual acfivity，and cervical carcinoma，"American Jouranl of Obstetrics and Gvnecology vol. 139(1981)pp. 52—57。

S. H. Swan and D. B. Petitti. "A review of problems of bias and confounding in epidemiologic studies of cervical neoplasia and oral contraceptive use，"American Journal of Epidemiology vol. 115(1982). pp. 10—18。

子宫颈癌死亡率一段时间来一直在下降；原因未知。这个例子是由 McGill 大学流行病教授 Michael Kramer 提出。

5. R. M. Movre et al. "The relationship of birthweight and intrauterine diagnostic ultrasound exposure，"Journal of obstetrics and Gynecology vol. 71

(1988)pp. 513－17. 混杂变量：种族，登记状况(公立或私立)，吸烟状况，分娩状况(足月或早产)，自发性流产史，饮酒史，羊膜穿刺状况，胎儿检查，分娩方法，教育，登记怀孕周数，胎儿期出诊次数，母亲的体重和体重增加。分娩时的怀孕期。

临床试验：U. Waldenstrom ef al. "Effects of routine one－stage ultrasound screening in pregnancy：a randomized controlled trial，"Lancet (Sept. 10，1988)pp. 585－88。平均来说，接受超声的婴儿比起对照者有较高的体重。在处理组，妇女们看见了她们所怀有的胎儿的超声像。结果，她们中的许多人放弃了吸烟，而吸烟会引起低出生体重的。吸烟习性的改变可能是保护效果的原因。

6. "Suicide and the Samaritans，"Lancet (Oct. 7. 1978)pp. 772－73(社论性的)。最早的研究人员是 C. Bagley (Social Science and Medicine，1968)。他没有将城镇与使用的可燃气类型相配，而这些数据似乎现在是不可能得到的。再制是由 B. Barraclorgh et al. (Lancet，1977；Psychological Medicine，1978)。我们发现这个例子是在 D. C. Hoaglin，R. J. Light，B. Mcpeek，F. Mosteller，and M. A. Stoto，Data for Decisions (University Press of America，1984，p. 133)。

7. 在 Berkeley 数据中的似非而又可能是的论点被 Eugene Hammel 注意到，即后来的研究生院的副院长。他在两位同事 P. Bickel 和 J. W. O'Connell 的帮助下解决了问题。我们沿着他们的报告，"Is There a sex bias in graduate admissions?"Science vol. 187(1975)pp. 398－404. 入学数据来自 1973 年秋季。

8. Statistical Abstract，1988，Table 117.

9. 见第 1 章注释 2。

10. 参考文献：

L. M. Friedman，C. D. Furberg，and D. L. DeMets Fundmentals of Clinical Trials ，2nd ed. (Littleton，Mass. ：PSG Publishers，1985).

T. L. Lewis，T. R. Karlowski，A. Z. Kapikian，J. M. Lynch，G. W. Shaffer，D. A. George and T. C. Chalmers，"A controlled clinical trial of aseorbie acid for the common cold ，"Annals of the New York Academy of Sciences vol . 258(1975)pp. 505－12。

T. R. Karlowski，T. C. Chalmers，L. D. Frenkel. A. Z. Kapikian，T. L.

Lewis, and J. M. Lynch, "Ascorbic acid for the common cold ," Journal of the American Medical Association vol. 231(1975)pp. 1038－42。

K. J. Carpenter, The History of Scurvy and Vitamin C(Cambridge University Press, 1986)。

11. "Nicotinic acid"是烟酸的专门名词，一种预防糙皮病的因素。显而易见地，引进术语"niacin"是因为"nicotinic acid"在面粉标签上看上去太不吉利的缘故。烟酸在冠心药物项目(the Coronary Drug Project)中被试验并且没有效果。习题中的表是虚构的。

12. 搭救在牲畜中坚持了许多年，其他的试验给出十分类似的结果。筛法使检测加快了大约若干个月，那样看来似乎足够有济于事。Johns Hopkins 大学流行病学荣誉教授 Sam Shapiro 友好地提供了未发表的资料。在 HIP 试验中，有一次初步的筛法检测和三次年度重筛，每一次包括由一名医生和早期胸部肿瘤 X 射线测定法进行的乳房检查。

参考文献：

J. Cairns, "The treatment of disease and the war against cancer," Scientific American vol 253(1985)pp. 51－59

S. Shapiro, "Evidence on screening for breast cancer from a randomized trial," Cancer vol. 39(1977)pp. 2772－82。

S. Shapiro etal. "Current results of hhe breast cancer screening randomized trial "在 N. E. Day and A. B. Miller, eds. Screenig for Breast Cancer (H. Huber, 1988)上。

L. Tabar et al. "Reduetion in mortality from breast cancer after mass screening with mammography," Lancet(April. 13, 1985)pp, 829－32。

13. 有关参考文献，见注释 4。

14. 这个例子由 California 州的卫生行政部门的 Shanna Swan 博士提出，基于来自在 Walnut Creek, California 的 Kaiser Permanente 做的观察研究的数据。

15. B. Rush, An Account of the Bilious Remitting Yellow Fever as it Appeared in the City of Philadelphia in 1793(Philadelphia: Dobson, 1794)。关于此流行病的现代历史，见 J. H. Powell, Bring Out Your Dead (University of Pennsylvania Press, 1949)。我们在 S. J. Pocock, Clinicol Trial (John Wiley & Sons, 1983)发现了引文。

美国独立宣言的签名者，Rush 是十八世纪美国最著名的医生之一。适度的放血和通便是无效的；而 Rush 只是改变为很大的剂量——同样的治疗。这些处理在十八世纪医学理论的环境中是讲得通的(Powell，第 3 章)。从现在的眼光，当对他的同时期的人作这样的治疗时他的疗程似乎是发疯似的。

16. Statistcal Abstract，1988，Table 268.

17. Federal Register，vol 53，No 224 Monday November 21，1988，p. 46971.

18. San Francisco Chronicle，December 9，1975。研究报告本身更加有节制：D. J. Ullyot et al. "Improved survival after coronary artery surgery in patients with extensive coronary artery disease。"Journal of Thoracic and Cardiovascular Surgery vol. 70(1975)pp. 405－13。

19. E. S. Pearson and J. Wishart，eds Student's Collected papers (Cambridge University Prees. 1942).

20. Statistical Abstract. 1971 Table 118. 研究在 1964 年进行。

21. 参考文献：

L. M. Friedman，C. D. Furberg，and D. L. DeMets，Fundamentals of Clinical Trials，2nd ed. (Littleton，Mass：PSG. Publishers ，1985)。

P. J. Schechter，W. T. Friedwald，D. A. Bronzert，M. S. Raff，and R. I. Henkin，"Idiopathic hypoguesia：a deseription of the syndrome and a single－blind study with zinc sulfate，"International Review of Neurobiology (1972)Supplement 1 pp. 125－39.

R. I. Henkin，P. J. Schechter，W. T. Friedwald，D. L. DeMets，and M. S. Raff，"A donble－blind study of the effecfs of zinc sulfate on tasfe and smell dysfunction，"American Journal of the Medical Science vol 272 (1976)pp，285－99。

22. 这个例子由 Shanna Swan 博士提出，见 E. Peritz et al. "The incidence of cervical cancer and duration of oral contraceptive use"，American Journal of Epidemiology vol. 106(1977)p. 462－69。在登记之前也对一些早期子宫颈痛涂片和"选择的传染病"作些调整。另外的参考文献见注释 4。

23. 在第 2 段中讨论的研究是由 Sheldon and Eleanor Glueck，Unraveling Juvenile Delinquency(Harvard University Press，1951. p. 120)所做。Kanl Bemesderfer 使它引起我们的注意。也见 T. Hirschi and H. C. Selvin，

Delinquency Research(New York:Free Press,1967,p. 79)。

第二部分　描述性统计

第3章　直方图

1. 由 Antoine de Saint Exupéry。经过出版者 Harcourt Brace Jovanovich. Inc. 的同意而再版。
2. Moneg Income in 1973 of Families and Persons in the United States Current Population Reports. Series P－60,No. 97(January,1975)U. S. Department of Commerce.
3. 1973 年的数据,见注释 2。1987 年的数据,见 Money Income of Households, Familids and Persons in the United States:1987. Current Population Report Series P－60,No162(February,1989)V. S. Department of Commerce.
4. 见注释 2
5. 对于小组区间这是精确的,对于其它区间则是近似的。
6. Statistical Abstract,1971,Table 118.
7. 许多变量可以归类为任一方式,依赖于你怎样看它们。例如,收入决不会相差少于一分钱。然而,将收入作为连续的量处理是方便的——因为它的跨程比起最小的变化来说是如此的大。
8. 用狭窄的小组区间,直方图可能会如此地外形参差不齐以致于它的形状不可能辨认出来。用较宽一些的小组区间,即使某些信息损失了,直方图的形状可以较易看出来。有关讨论,见 P. Diaconis and D. Freedman,"On the histogram as a density estimator: Lz theory." Z. Wahrscheinlichkeitstheorie vol. 57(1981)pp. 453－76.
9. 见 I. R. Fisch,S. H. Freedman,and A. V. Myatt,"Oral contraceptive,pregnancy,and blood pressure,"Journal of the American Medical Association vol. 222(1972)pp. 1507－10。我们的讨论是因这篇文章而起,我们感谢 Shanna Swan 博士与 Michael Grossman 博士的技术性忠告。

   血压取为两种状态,收缩的和舒张的。我们考虑收缩状态。关于舒张状态的结果十分类似。血压是通过比较水银柱施加的压力而测量的:它用水银柱的表度来表示——因而单位是"mm"或者毫米水银柱。在避孕药物研究中,血压是用机器测量的。研究排斥了大约 3500 名妇女,她们是怀孕的,产后的,或者服用不同于口服避孕药的激素药物;这些因素影响血压。药物研

究课题发现4个年令组是足够的：17－24，25－34，35－44和45－58。在这些年令组的每一组内使用者与不使用者的年令分布十分类似。

10. R. C. Tryon，"Genetic differences in maze－learning techniques in rats." 39 th yearbook，National Society for the Study of Education part I (1940)pp. 111－19. 这篇文章在一本非常好的读物上再版：Anne Anastasi. Individual Differences(John Wiley & Sons，1965)。Tryon对他的直方图使用了一非线性尺度，因此它们看上去十分不同于我们的草图。

11. 1970 Census of Population vol，I part 1 Section 2 Appendix，p. 14 U. S. Department of Commerce。对于1880年这一列仅列入了23—99岁的人；对于1970年这一列仅列入了23—82岁的人。

12. K. Bemesderfer and J. D. May，Social and Political Inquiry (Belmont. Calif：Duxbury Press，1972 p. 6)。

13. **参考文献是**：

R. A. Baron and V. M. Ransberger，"Ambient temperature and the occurrence of collective violence：the 'long ，hot summer 'revisited，"Journal of personality and Social psychology vol. 36(1978)pp. 351－60 引文轻微地校订过。

J. M. Carlsmith and C. A. Anderson，"Ambient temperature and th occurrence of collective violence：a new analysis，"Journal of Personality and Social Psychology vol. 37(1979)pp. 337－44.

图是根据Baron与Ransberger中原物再绘制的，那是the American Psychological Association的版权：经得版权所有者和作者的允许后再版的。

第4章 平均数和标准差

1. Natural Inheritance (London；Macmillan，1889；由the American Mathematical Society Prees. 1973再版。)

2. 直方图最高的点，众数，有时侯用来指示中心。这不作推荐，因为数据中的微小变化可以引起众数的较大移动。

3. T. Alexander，"A revolution called plate tectonics"Smithsonian Magazine，vol. 5. no 10(1925). A Hallam，"Alfred Wegener and the hypothesis of continental drift."Scientific American vol. 232 No. 2(1975) Ursula Marvin，Continental Drift(Smithsonian Institution Press，1973)。

4. **参考文献是**Blood Lead Levels for persons Ages 6 Months－74 Years：Unit-

ed states,1976－1980。数据来自 the National Health Survey,Series 11,No. 233,U.S. Department of Health and Human Services,Washington,D.C. 公共卫生总署和全国卫生统计中心设在卫生与公众服务部内。下面例子的数据取自人口与卫生统计出版物的 Series11,以及来自全国卫生统计中心和政策与社会研究的校际合作所提供的数据带。我们对这些数据的所有解释,不管正确或者错误均负责。我们对 Arthur J. MeDowell 先生,1976 年健康检查统计部门的负责人,在第一版中的帮助表示感谢。我们感谢 Dorothy Rice 教授(University of California,San Francisco)和 Dale Hitehcock 教授(NCHS)对于第二版的帮助。

在图 4,8,和 9 中的直方图基于不加权的样本计数;同样地概括统计量也是如此,它是经过四舍五入的,加权几乎不产生差异。确切数值(平均数±SD)如下:

	男子 18－74 不加权	男子 18－74 加权
身高	68.78±2.83	69.11±2.82
体重	170.92±30.13	172.19±29.75
收缩压	131.55±19.29	128.91±17.70
舒张压	81.88±17.96	81.52±17.92
	妇女 18－74 不加权	妇女 18－74 加权
身高	63.46±2.62	63.71±2.60
体重	145.71±32.65	144.18±32.27
收缩压	127.37±23.42	122.81±21.22
舒张压	78.40±18.53	77.32±18.31

5. 图 3 中的组:18－24,25－34,25－44,45－54,55－64,65－74。见 Anthropometric Reference Data and Prevalence of Overweight:United States,1976－1980。数据来自全国健康调查,Vital and Health Statistics,Series 11,No 238,U.S. Department of Health and Human Services,Washington,D.C.

长期趋势被估计为每十年大约 0.4 英寸;在 1960－1980 年整个 20 年期间,平均来说美国人高 0.8 英寸。此外,人们从 50 岁到 75 岁间看来身高少了 0.5－1.5 英寸。(一种可能的解释:大约 2 英寸的身高是由体内骨骼之间的空隙所形成;随着年令的增加身体自身压紧,因此这些空隙越来越

少。)长期趋势与缩小提示了从 20 岁到 70 岁总的下降 2.5—3.5 英寸。观察到的下降男子为 2.3 英吋而妇女为 2.1 英寸，因此可能有其它的因素在起作用。我们乐于感谢 Reubin Andres 博士(NIH)和 Stanleg Garn 博士(University of Michigan)的帮助。

关于身高趋势作为社会变化的象征的讨论，见 R. Floud，K. Wachter，and A. Gregory，Height，Health and History(Cambridge University Press，1991)。

6. 对于整数数据和中心在整数的小组区间这是精确的。在其它状态下，它仅仅是近似。
7. 见第 3 章的注释 3。
8. 基本理由是由统计学家所称的正交性。简略地，当某些场合的误差由一些独立的来源所引起时，有一个简单的和精确的公式以得到总误差的 r.m.s. 大小：r.m.s. 误差象直角三角形的边一样结合在一起。具有两个正交的误差来源，

$$C=\sqrt{a^2+b^2}$$

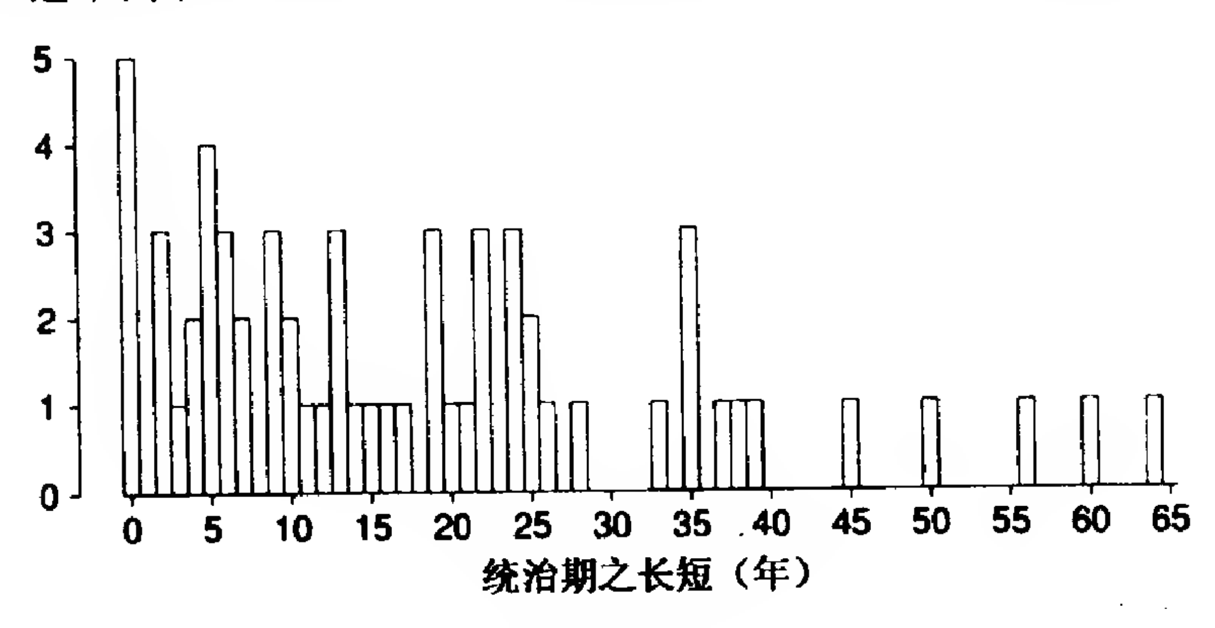
其中 a 是来自一来源的误差的 r.m.s. 大小，b 是来自另一来源的误差的 r.m.s. 大小，而 c 是总误差的 r.m.s. 大小。这个事实在本书中将若干次用到：在回归(第三部分)，在计算的和标准误差(第五部分)，以及在计算差异的标准误差(第八部分)中。使人惊奇地，没有这样的公式可用于平均绝对值。

9. 数字是经过四舍五入的，概括统计量的精确值记录在注释 4 中。有关身高数据在范围之内确切的百分数如下所示。

	平均数±1SD	平均数±2SD	平均数±3SD
男子	68.1%	95.1%	99.6%
妇女	68.3%	95.2%	99.6%

即使对不遵循正态曲线的许多数据组，68%—95%准则运转得十分好。取乔治六世整个时期 61 名英国君主的统治期长短为例。这些时期平均数为 18.1 年，具有 15.5 年的 SD。它们的直方图如下所示，它没有一点象正态曲线，仍然 61 个内有 42 个，或者 69%，在平均数的 1SD 以内。61 个中的 57 个，或者 93%，在平均数的 2SD 以内。(由定义，一个统治期的长短是其第一年与

最后一年之间的差，鉴于 Information Please Almanac 1988 年的 pp. 274－75 的报告；这个例子是由 Minnesota 大学统计系的 David Lane 提供的）



10. SD 的平方数称为方差。这常常用来作为散布的一个测度，可是我们不推荐它作为一个描述性统计量。例如，关于美国男子其体重的 SD 大约是 30 磅：粗略地说来单独男子离平均体重 30 磅。体重的方差是

$$(30\text{ 磅})^2 = 900\text{ 平方磅。}$$

11. 然而，这个公式对于舍入误差是敏感的。

12. 见注释 5 作为参考：数字已经精简。

第 5 章　数据的正态近似

1. 也称作标准分，Z－分，西格玛一分。

2. Data on college－bound seniors from a College Board Press release, September 20, 1988，由 Paul Holland（ETS）友好地提供。关于 SAT 分下降的讨论，见 Willard Wirtz et al. On Further Examination: Report of the Advisory Panel on the Seholastic Aptitude Test Score Decline (New York: College Entance Examination Board, 1977)。

第 6 章　测量误差

1. NBS 已更名为 NIOST—the National Institute of Science and Technology（全国科学技术协会）。我们乐于感谢标准局 H. H. Ku 博士的帮助。

2. 重量在这里使用以代替更技术性词质量 1983 年，the General Conference on Weights and Measures 代替 the Treaty of the Meter（米突公约），但是重量仍然通过参照标准物来确定。顺便提一句，在巴黎的国际原型千克似乎在减少重量且没有人知道为什么原因。见"Move over, Oprah: Now even weights are losing weight," Wall Street Journal (July12, 1990) p. A1.

3. 在标准局的精确称量中机会误差的两个较大来源被认为是：

 · 在平衡装置特别是(天平的)刃形支承上游动的微小量；

 · 砝码在平衡秤盘中的位置中的轻微变异。

4. P. E. Pontius,"Measurement philosophy of the pilot program for mass calibration,"NBS Technical Note No. 288(1966)。标准局拒绝异常数据仅仅是"因为诸如门砰地开关或设备失灵这样的原因。"

5. J. N. Morris and J. A. Hardy,"Physique of London busmen,"Lancet(1956) pp. 569－70。这篇参考文献是由 Jerusalem(耶路撒冷)的 Hebrew 大学统计系 Eric Peritz 教授提供。

6. 观察的数据在 H. Zeisel,H. Kalven,Jr,and B. Buchholz,Delay in the Court,4 th ed. (Little,Brown & Co,1959)中讨论。实验在 Maurice Rosenberg,The Pretrial Conference and Effective Justice(Columbia University Prees,1964)中报导。定义在 p. 19 上,数在 p. 20 上,以及数据在 pp. 48 与 52 的表 8 与 9 中。关于另一讨论,见 H. Zeisel,Say It with Figures,6 the ed. (Harper & Row,1985,p. 141)。

7. 见第 3 章的注释 3。

8. 见第 4 章的注释 5。

第三部分 相关与回归

第 8 章 相关

1. 有处理多于两个变量的方法,但是这些方法十分复杂。需要一些矩阵代数知识以进行讨论。参考文献是:

 C. Daniel,F. S. Wood,and J. W. Gorman,Fitting Equations to Data,2nd ed. (John Wiley & Sons,1980);

 N. R. Draper and H. Smith,Applied Regression Analysis,2nd ed. (John Wiley & Sons,1981);

 C. R. Rao Linear statistical Inference and Its,Applications,2nd ed(John wiley & Sons,1973)

 H. Scheffé,The Analysis of Variance(John Wileg & Sons,1961)。

 在第 12 章第 3 节有一个多元回归的简短讨论。

2. K. Pearson and A. Lee, "On the laws of inheritance in man," Biometrika, part ii(1903)pp. 357－462,给出了联合分布,其中身高舍入到最近的英寸。我们加上始终如一的噪声以得到连续的数据。归因于独立的随机化,这里的数据集有点儿不同于我们第一版中的东西。

3. 这个术语不是标准的。

4. H. N. Newman, F. N. Freeman, and K. J. Holzinger, Twins: A Study of Heredity and Environment(University of Chicago Press, 1937)。在双胞胎研究中,惯例是每一对双胞胎标两次:一次为(x,y),一次为(y,x)。

5. 数据来自于 1988 年三月的现场人口调查;相关系数是从由人口普查局提供的数据磁带而计算,并且是不加权的。收入来自 1987。

6. 见注释 5。"孩子数"是拥有的 18 岁以下居住在家的孩子数。

7. 当相关为零时,任一斜率可以使用。"SD 线"不是标准术语。

8. 然而,这个公式可以招致舍入误差。

9. Consumption Patterns of Household Vehicles, 1985. Residential Transportation Energy consumption Survey. Energy Information Administration, Washington, D. C.

10. Marjorie Honzik 博士(Institute of Human Development, Berkeley)相当友好地提供该数据。

第 9 章　再谈相关

1. NOAA, Climatological Data(Asheville, North Carolina, 1985)。

2. R. Doll, "Etiology of lung cancer," Advances in Cancer Research vol. 3 (1955) pp. 1－50. Report of the U. S. Surgeon General, Smoking and Health(Washington, D. C, 1964)。

3. 该想法追溯到 W. S. Robinson, "Ecological correlations and the behavior of individuals," American Sociological Review vol. 15 (1950) pp. 351－57. Robinson 基于 1930 年人口调查数据给出了识字与种族的例子。我们的例子是一个重复,使用了由人口普查局提供的有关 1988 年三月现场人口调查的数据带。

如果每一群是二元正态的,具有公共的回归线,那么斜率和截距可以从平均数估得。也见 L. Goodman, "Ecologlcal regression and the behavior of individuals," American Sociological Revieaw vol. 18(1953)pp. 663－64。关于在法律情况中的一个例子的批判性评注,见 D. Freedman, S. Klein, J.

Sacks et al "Ecological regression and voting rights," Technical Report No. 248, Statistics Department, University of California Berkeley, Evaluation Reiview vol. 15(1991)

4. E. Darkheim, Suicide (Macmillan, 1951, p. 164)。我们计算了相关系数。Durkheim 考虑了省份群的平均数，该情况相关系数为 0.9：他的结论是"公共教育与自杀是同分布的。"

5. 多元回归有点帮助，但是比起它回答的，常常出现了更多的问题(第 12 章，第 3 节)。

6. 有关更多的讨论，见 H. Zeisel. Say It with Figures, 6th ed. (Harper & Row, 1985 pp. 152ff)。

7. 数据由 M. Russell 提供，来自于 D. Jablonski, "Larval ecology and macroevolution in marine invertebrates," Bulletin of Marine Science vol. 39 part 2(1986)pp. 565－87 中的表 1。也见 Science vol. 240(1988)p. 969。

8. R. Doll and R. Peto, The Causes of Cancer (Oxford University Press, 1981)。有一个来自于流行病和动物实验的强烈的证据表明吃得过多是致癌的。(在等热饮食中)脂肪的效果是几乎不清楚的，也看 A. Schatzkin et al. "Serum cholesterol and cancer in the NHANES I epidemiologic followup study, Lancet ii(1987)pp. 298－301；此研究提出了来自脂肪的保护效果！最近的预测性研究支持了该生态分析。见 W. C. Willett et al. "Relation of meat, fat and and fiber intake to the risk of colon cancer in a prospective study among women" New England Journal of Medicine Dec. 13, 1990, pp. 1664－71.

9. National Assessment of Educational Progress The Reading Report Card (Princeton: ETS/NAEP, 1985, p. 53)。

10. T. R. Dawber et al. "Coffee and cardiovascular disease: observations from The Framingham study," New England Journal of Medicine vol. 291 (1974)pp. 871－74。

11. M. P. Rogin and J. L. Shover, Political Change in California (Westport, Conn: Greenwood Press, 1970, P. xvii)。

12. 数据来自 1988 年三月现场人口调查；数据磁带由人口普查局提供。

13. 这重复了由 M. 和 B. Rodin 的研究，"Student evaluations of teachers," Science vol. 177(1972)pp. 1164－66. 以单独的水平，相关将是较弱的；然

而，这是一个引起兴趣的迹象。

14. D. Bahry and B. D. Silver,“Intimidation and the symbolic uses of terror in the USSR”American Political Science Reviewnw vol. 81(1987)pp. 1067—98 研究人员同时调整了某些其它可能的根据调查表测定的混杂变量：物资满意度，政权准则的支持，人民的信仰，对政策的兴趣，与 KGB 相处的以前的经验。

15. T. W. Teasdale et al.“Degree of myopia in relation to intelligence and educational level ”Lancet(Decemher 10,1988)pp. 1351—54

16. 数据来自 College Board press release,September20,1988，由 Paul Holland (ETS)友好地提供。50 个州平均 SAT 分的平均数是 498,SD 是 32。50 个州百分率的平均数是 34，具 SD 为 25。纽约百分率是 72；怀俄明州，12。

第 10 章　回归

1. 这些数字有点舍入。确切的数(未加权的)是：

   平均身高＝69.6 英寸　　SD＝2.79 英寸

   平均体重＝162 磅　　SD＝29.1 磅　　r＝0.448

   数据来自政治和社会研究的校际合作所提供的数据磁带。

2. 数据来自 1988 年三月现场人口调查；统计量计算出自人口普查局提供的数据磁带，未加权的。

3. 术语“平均曲线”不是标准的。大体上，曲线依赖于 X 值被细分的程度。

4. 专用术语没有很好地标准化；有时候，第 1 个百分序是最好的。

5. 父亲的平均身高是 67.6 寸英，有 2.74 英寸的 SD；儿子平均身高是 68.7 英寸，有 2.81 英寸的 SD；r 为 0.501。从最初的交叉表，SD 是 2.72 英寸和 2.75 英寸，r≈0.514；四舍五入对平均数只有很小影响。

6. Marjorie Honzik 博士提供了数据。作为比较，在 1988 年三月现场人口调查中，丈夫与妻子的教育水平之间的相关系数是 0.64。

第 11 章　回归的 R. M. S. 误差

1. Statistical Methods for Research Workers(Oliver &. Boyd,1958,p. 182)。

2. 在多元回归中，残差可以对因变量，每一个自变量，拟合值，和省略的变量来点图。

3. 有 60 个家庭其中父亲高 64 英寸(舍入到最近的英寸)；儿子身高平均为 66.7 英寸，具有 SD 为 2.29 英寸。有 50 个家庭其中父亲高 72 英寸(舍入到最近的英寸)；儿子身高平均为 70.7 英寸，具有 2.30 英寸的 SD。

4. 在本书的这部分，中心主要是描述性的。有一组有限数据，每一点赋予相等的权。回归线看作使 E{Y|X}在这个有限总体里光滑化。对于给定的总体中的一个随机元素 Y 从 X 得到预测。$\sqrt{1-r^2}$公式作为预测误差在这个模型中是正确的。

如果回归线用来拟合训练样本并且然后用来作出预报，那么如同上面讨论的那样在回归线附近存在一个方差分量。样本回归线也有一个方差分量在总体线附近。后者当然依赖于 X，这超出了本书的范畴。但是，带有一个大的训练样本，则第二个方差分量可能相当小。以第 5 节中例 1 为例，它拥有大小为 100 的训练样本。第一个方差分量(在回归线附近)大约是 64。对于一个离 LSAT 平均数 z SD 远的学生，第二个方差分量——归因于回归线中的抽样误差——在约是 $0.64(1+z^2)$。一个相当极端的情况是 z=3。总 r.m.s 预测误差大约是 8.4；不顾第二个分量则给出 8.0。

5. 数据来自 1988 年三月现场人口调查；统计量计算出自由人口普查局提供的未加权及四舍五入过的数据带。收入来自 1987 年。包括任何劳动力状态的妇女在内。每一点可能表示若干名妇女，尤其是零收入的妇女。

6. 在 Mead 湖的三号站；获得数据的 53 个月的几何平均数；一年到头；抽样时间 1976—1986；Jerome Horowitz(Las Vegas)提供的数据和一次关于水质标准的听证会有关。

7. 数据由 Marjorie Honzik 博士提供。

8. 投手排斥在外；但是，第二年的消沉也可以根据投手的平均得分而观察到。1979 年，在 the American League 有二个"当年的新星"；1975 年在 National League 有二个——两次均为投手。数据由 STATS，Lincolnwood，Illinois 所提供。David Lane 提供了关于棒球统计的专家指导。

第 12 章 回归线

1. Abhandlungen zur Methode der Kleiusten Quadrate(Berlin，1887，p. 6)。我们采用由 L. LeCam and J. Neyman 的翻译本，Bayes—Bernoulli—Laplace Seminar(Springer—Verlag，1965，p. viii)。

2. 数据来自 1988 年三月现场人口调查；从由人口普查局提供的数据带中计算统计量，不加权与四舍五入的。收入来自 1987 年，不论劳动力状态。每一点可能代表若干人，尤其是具有零收入的点。

3. 参见注 2。

4. 这个等式是基于由 IRRI 提供的概括统计量的四舍五入后的值。

5. 参见注 2。

6. 这些统计量的计算出于由 Inter－University Consortivm for political and Social Research 提供的数据带。样本容量是 1065，因此斜率是实际的，而不是抽样中的偶然性。(删失数据的情况排斥在外：概括统计量是经过四舍五入的。) 也可参见 T. W. Teasdale et al. "Fall in the association of height with intelligence and educational level," British Medical Journal vol. 298 (1989)pp. 1292－93。

7. 由 Berkeley 物理系的教授 William Fretter 实施，作为他的初等物理教程中的一个示范。

8. 在 Berksons 的误差变量模型中这里回归是恰如其分的。砝码名义上的值由研究人员确定，真实的值则有受误差影响；名义值进入回归。当称量砝码的值遇到误差时，测量值进入回归，于是通常的回归估计是有偏的。参考文献是 G. W. Snedecor and W. G. Cochran, Statistical Methods, 6th ed (Iowa state University Press, 1973)。

9. 关于社会科学中回归模型的更多的讨论，见 1987 年夏季写的 Journal of Educational statistics 也见 1987－1991 年的 sociological Methodology

10. 见注释 6

11. "盐中：电解质分泌和血压的一个国际研究。24 小时的尿中钠和钾的分泌结果，" British Medical Journal vol. 297(1988)pp, 319－28。作者表明了盐的摄取和年龄之间的相互作用。关于明确的实验结果一在低摄入量一见"轻微高血压随着饮食盐摄入量的适度减少血压会下降。" Lancet (February 25, 1989)pp. 309－402。

12. 由 Franklin Fisher (MIT) 在 Cuomo v Baldrige (80 civ 4550SDNY 1987) 中给出的关于调整调查(或统计数字)的回归模型的证明的释义，改写本 pp. 2149ff.

第四部分. 概率

第 13 章. 机会是什么?

1. 有关机会的其它观点参见：

R. A. Fisher. Statistical Methods and Scientific Inference, 3rd ed.

(New York: Hafner,1973);

L. J. Savage, Foundations of Statistics, 2nd ed. (Dover, 1972)。

2. 第三版在 1756 年,de Moivre 死后出版。由 Chelsea Publishing, New York, 1967,再印。

3. Statistical Abstract, 1988, Table13.

4. W. Fairley and F. Mosteller。"A Conversation about Collins," University of Chicago Law Review(1974)。

5. 检察官对两个"事件"计算了两个"机会",在它们之间来来回回地考虑。第一个事件是被告有罪的。第二个事件是在洛杉矶没有一对与描述相符。对于一个频率论者,机会概念应用得不怎么样好。即使是一些 Bayes 统计学家在这里可能感到一些困难,因为没有合适的机会模型把有罪或者无罪的假设与数据联系起来。

   在洛杉矶存在其它的对与描述相符吗? 首先,这似乎可能象一个通过取一组样本而得到解决的统计问题。但是,计算将证明该城市中一对对地抽样不能以适当的置信水平解决问题:需要完全的人口普查。

6. 本习题由 D. Kahneman and A. Tversky, "Judgement vnder uncertainty: heuristics and bias," Science vol, 185(1974)pp. 1124—31,提出。也见 D. Kahneman, P. Slovic, and A. Tversky, eds。Judgement under Uncertainty: Heuristics and Bias(Cambridge University Press, 1982)。

第 14 章. 再谈机会

1. 选自 The Doctrine of Chances 的献辞。

2. 在 Pascal 和 Fermat 之间通信的有关另外历史记载,见 F. N. David, Games, Gods and Gambling(Buckinghamshire, england: Charles Griffin & Co, 1962)pp. 88--89。

第 15 章. 二项系数

1. 明显地,Jia Xian 发现二项系数略早一些——在十一世纪。见 Li yan 和 Du Shiran, Chinese Mathematics(Orford University Press, 1987, pp, 121ff。) Pascals 三角形称作杨辉三角形,也许更公正些。

2. 模型仅仅是近似的;甚至比新生儿是男性的机会略好一点,家庭中的连续出生孩子稍微有点相依。

3. 习题由 D. Kahneman and A. Tversky。"Judgment under uncertainty: heuristics and bias," Science Vol。185(1974)pp. 1124—31 提出。

第五部分. 机会变异

第 16 章. 平均数律

1. An Experimental Introduction To the Theory of Probability(University of Witwatersrand Press。1964)。二次大战以后 Kerrich 在南非执教。
2. 我们不谈及"绝对值",即将符号省略的值。代之以,我们试图区分此数(在绝对意义来说)的差异和以百分比的差异。
3. 我们在非技术性意义下使用了"机会过程"一词,一个"由机会过程所产生的数"是随机变量的观察值。
4. 计算程序是定数性质的,因此不可能产生真正随机形式的数。但是,程序可以产生一系列似乎十分随机的数。一个方法是乘以一个相当大的数 M。由程序编制人员选择"种子"x:x 在 0 与 1 之间。计算机作出 M 乘以 x,得到一个整数部分和一个小数部分:

   …aaaaaaaaaaaa · bbbbbbbbbbbbb…

   小数点左面的数字打印出来作为第一个随机数,而小数部分用来作为下一个随机数的种子。进一步的讨论。见

   D. Knuth,The Art of Computer Programmimg Vol. 11,2nd ed,(Addison-Wesley,1981)。

   G. Marsaglia,"Random numbers fall mainly in the Planes。"Proceedings of the National Academy of Science Vol. 60(1968)pp. 25—28——,"A Current view of random mumder generators,"Keynote address,Sixteenth Symposium on the Iuterface betueen Computer Science and Statistics,1984。
5. "从盒中抽得数之和"不是一个标准术语但比起"独立同分布的随机变量之和"较松散一些。"盒子模型"也不是标准的,尽管它似乎能被理解。

第 17 章. 期望值与标准误差。

1. Keno 是(Las Vegas)拉斯维加斯对应于 bingo 的一种赌博。有 80 只球,编号从 1 到 80。每一次赌,随机不放回地选择 20 只球。例如,如果你就单个数 17 打赌,那末你就是在赌编号 17 的球将在这选出的 20 只球中间。
2. 本书中,我们对数据使用 SD 而对机会量(随机变量)使用 SE。区别不是标

准的，术语 SD 常用于这两者的情况。

3. E. O. Thorp, Beat the Dealer (Random House, 1966)。在 baccarat (一种纸牌游戏)的某些超过一般赌注的赌博也有正的期望值。对于跑道上最有希望获胜的马(令人难以理解)的赌博也是这样的：见 R. T. Thaler and W. T. Ziemba, "Parimutuel betting markets: racetracks and lotteries," Journal of Economic Perspectives Vol. 2(1988) pp. 161—74。

第 18 章. 概率直方图的正态近似

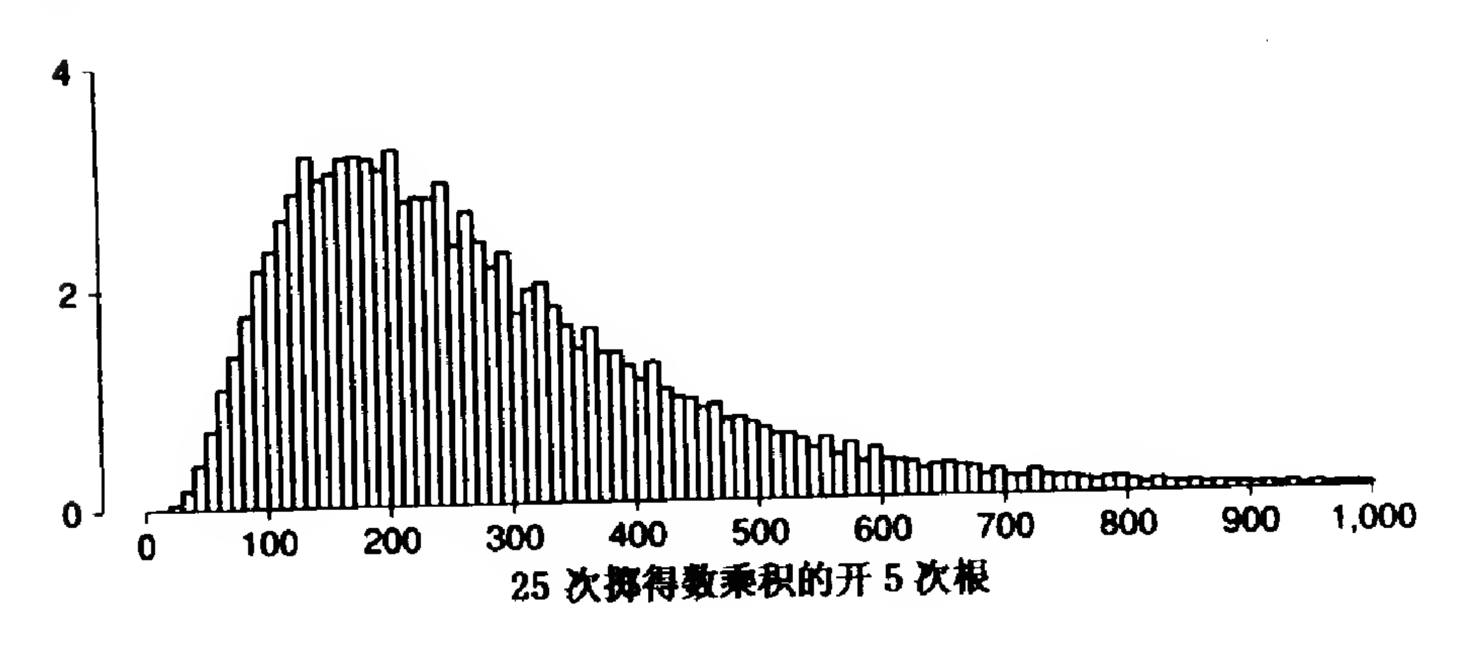
1. 在 Hewlett Packard HP 15C 上计算。
2. 数学讨论可以在 W. Feller, An Introduction to Probability Theory and its Applications vol. I. 3rd ed (John Wiley & Sons, 1968)的第 7 章中找到。
3. 如果抽取次数非常大，将数值的范围归在一起也许是有帮助的，如同图 10 (关于骰子掷出数的乘积)那样。如果票子没有显示全部数字也同样处理。即使抽取数是中等大小，并且票上显示全部数字，使用较宽的小组区间可能是有益的；当票上的数之间的差有大于 1 的公约数时就是这样。例如，假使你在抛硬币中赌 1 美元，共进行 100 次，你的净利就好象从盒子 [-1] [+1] 中作 100 次随机放回的抽取所得数之和，净利的可能值是偶数：0，±2，±4，…以这些值为中心宽为 1 的矩形可以绘制直方图。但是，采用第 4 节的方法，宽为 2 的矩形产生更好的结果。本书中的问题不产生这类问题。
4. 在 HP 15C 上计算而得。
5. 该偏差影响的数学分析由 Edgeworth 展开提供：在 W. Feller, An Introduction to Probability Theory and its Applications vol. II, 2nd ed. (John Wiley & Sons, 1970)中的第 16 章中有一个讨论。
6. 波浪可以这样来解释：如果盒子是 [1] [1] [9]，和的可能值将是 25，33，41，…被量为 8 的间距而分隔。如果盒子是 [2] [2] [9]，和的可能值将是 50，57，64，…被量为 7 的间距而分隔。图 9 中的盒处于这两个盒子的中间状态，因而峰与峰间的距离在 7 与 8 之间交替呈现。另一种考虑途径：波峰反映了在 25 次抽取中数字 9 的分布。
7. 图 10 中直方图的形状可能有点儿惊讶。但是，如果 $x_1, x_2, \cdots$ 是骰子的一连串掷得数，那末在再标刻度之后

$$(x_1x_2\cdots x_n)y^{1/\sqrt{n}}$$

近似地为对数正态。25 次掷得数乘积的开 5 次根的概率直方图在下面给出，它有正确的形状。概率是使用组合算法求得，摆动是真实的。(一颗骰子的 25 次掷得数的乘积具有形式 $2^a3^b5^c$。a，b，c 为非负整数；这个形式有助于间隔和摆动 。)

25 次掷得数乘积(以 10 为底)的对数是 25 个对数之和。每一个对数具有平均数 0.4762 与 SD 0.2627，因此 25 个对数之和具有期望值 $25\times0.4762\approx11.91$ 和标准误差 $\sqrt{25}\times0.2627\approx1.31$。25 个对数之和已经相当接近于正态地分布。

取图 10 中关于 25 次掷得数之乘积的最下面部分。轴在 10^{13} 处中止，这在对数刻度上是 13，或 0.83 个标准单位。大约 20% 的概率在该值的右边。在直方图中每个矩形的宽为 10^{11}。第 1 个矩形覆盖了从 $-\infty$ 到 11 的对数刻度区间，以标准区间是 $(-\infty,-0.69)$。这个区间包含了概率的大约 25%！



8. 默认的假设：非零 SD，和在盒内有限张票。
9. 假设两个盒内的票有同样的平均值，和关于平均数有同样的平均绝对偏离。如果它们也有同样的 SD，那末和的渐近习性将是相同的。如果 SD 不同，则渐近习性不相同。一个例子为

A) $\boxed{-1}\boxed{+1}$　　B) $\boxed{-2}\boxed{0}\boxed{0}\boxed{2}$

在这两个盒内，票的平均数是 0，关于平均数的平均绝对偏离是 1。但是盒 A 的 SD 是 1，而盒 B 的 SD 大约是 1.4：因此，(用任何合理的散布测度)来

自盒B中100次抽取之和的散布是来自盒A中100次抽取之和的散布的1.4倍。控制抽得数之和的渐近分布的是盒内数的平均数及SD，不是位置与散布的其他测度。

10. 设n表示抽取次数，k为重复次数。隐含的条件是$k/\sqrt{n}\log n \to \infty$。参见D. A. Freedman,"A Central Iimit theorem for empirieal histograms,"Z. Wahrscheinlichkeitstheorie Vlo. 41(1977)pp. 1—11。

第六部分. 抽样

第19章. 抽样调查

1. 本章开头的引语摘自The Adventure of the Copper Beeches. 我们在Don McNeill, Interactive Data Analysis (John Wiley & Sons, 1977)中发现了它。
2. 有关抽样的若干参考文献：

较少技术性的。

N. M. Bradburn and S. Sudman, Polls and Surveys (Jossey-Bass Inc, 1988)。

A. Campbell, G. Gurin, and W. Miller, The Voter Decides (Evanston: Row, Peterson, 1954).

Jean M. Converse, Survey Research in the United States: Roots and Emergence, 1890—1960 (Berkley: University of California Press, 1987)。George Gallup, The Sophisticated Poll Watcher' s Guide (Princeton Opinion Press, 1972)。

Herbert Hyman and others, Interviewsing in Social Research (University of Chicago Press, 1954)。

Frederick Mosteller and others, The Pre-Election Polls of 1948 (New York: Social Science Research Council, 1949)。

Mildred Parten, Surveys, Polls and Samples (Harper & Row, 1950)。

F. F. Stephan and P. J. McCarthy, Sampling Opinions (John Wiley & Sons, 1958)。

较多技术性的。

W. G. Cochran, Sampling Technigues, 3rd ed. (John Wiley & Sons,

1977)。

M. H. Hansen, W. N. Hurwitz, and W. G. Madow. Sample Survey Methods and Theory (John Wiley & Sons, 1953)。

Leslie Kish, Statistical Design for Research (John Wiley & Sons, 1987)。

3. 所有的引语都摘自 The New York Times (Oot ,1—15,1936)。

4. 我们要感谢盖洛普(Gallup)公司的 Diane Colasanto, Laura Kalb, Jack Ludnig, Coleen McMurrary, 和 Paul Perry, 他们回答了我们关于信息的所有要求。图 3,4 和 5 是依据盖洛普公司提供的复本而排版,并且得到他们友好的允许而再制。

5. 参见 Parten, P. 393, 及 Stephan and McCarthy pp. 241—70(note 2).

6. 另一种讨论,参见 M. C. Bryson, "The Literary Digest Poll: making of a statistical myth," American Statistician, Vol. 30(1976)pp. 184—85。Bryson 同意《文字摘要》民意测验被不回答偏差搞糟了。但是,他对选择偏差成为问题这一点持怀疑态度,并且怀疑 The Digest (文字摘要)是否真正地根据电话号码作为它的邮寄名单(即,民意测验被访问者的名单)进行抽样。

我们关于 The Digest 民意测验的信息的主要来源是 George Gallup—— 一个精明的、有(利害)关系的,和直接的观察者。他强调 The Digest 使用电话簿,汽车拥有者名单,和它自己的订购名单作为邮寄名单的原始依据。这种重要性,至少在实质上,被其他一些人象 Parten 或者 Stephan and McCarthy(参见注 2)所证实。

The Digest 没有发表我们能找到的有关它的过程的任何非常充分的论述。然而,通过对 1936 年 8 月 22 日至 11 月 14 日期间的 The Digest 期刊的复阅可以获知一些东西。例如,在 8 月 22 日一期的第 3 页上,我们发现:

> 民意测验描绘了三十年不断的进展与改善。基于使用了一个多世纪的由出版社推销书的"商业抽样"方法,现有的邮寄名单来自于美国的每一本电话簿,来自于俱乐部和协会的花名册,来自于城市姓名地址录,登记选民的名单,分类的邮寄订单和业务数据。[花折线部分是我们划上的]

文章继续解释,为了 1924 年的选举将名单放在一起,但随后由"训练有素的专家"修改。名单中的大多数姓名年复一年地继续使用,名单被用

于选举之间的民意测验(8 月 29 日期刊的第 6 页)。从登记选民的名单中抽取似乎于 1936 年是个新方法,而这样的名单仅仅对某些"大城市"采用(10 月 17 日期刊第 7 页。):

按现代标准,Digest 的邮寄名单以稍微任意的方式放在一起,它是有偏性的:它排斥了社会中大量的,可看作相同的部分。Bryson 暗示如果 Digest 的某种方式设法从它的 1 千万名单中得到 100%的回答,将可能预测选举结果。这似乎是不可能的,尤其因为 Digest 希望得到正确的百分数(8 月 22 日期刊第 3 页):

再一次地,Digest 访问一千万以上的选民——每 4 个中有 1 个,代表了美国的每一个县——以在十月份确定 11 月份的选举。

第二个星期,来自于这一千万选民的第一批回答开始了投票行情的涨潮,选票经过3 次检查,核实,5 次交叉核实并计算总数。当最后的数字经过合计与核查,如果以过去的经验为准则,那末选民将在 1 个百分数范围内知道四千万人中真实的公众投票。[花折线部分是 Digest 划上的。]

The Digest 相差了 19 个百分数点。正如我们在课文中所说的,有两个主要的原因:选择偏差和不回答偏差。

7. 这个 65%是 80 年代后期有代表性的 4 次呼唤概率抽样。回答率从 1975 年的大约 75%和 1960 年的 85%下降了。这种下降趋势是民意测验机构的一个主要担忧。

8. 本节吸收了 Mosteller 与其它人(参见注 2)的书的内容。

9. Stephan and McCarthy,p. 286(注 2)。

10. 混淆定额抽样与分层抽样是吸引人的,但两者是不同的。例如,假设想在对性别进行控制下从某镇抽取容量为 200 的样本;事实上,使得男性数等于女性数。定额抽样者原则上可以雇用两名访问人员,一名访问 100 位男性,另一名访问 100 位女性。在其它方面,两名访问人员将挑选他们所想的任何人。这不是一个好的设计。相比之下,分层抽样将如下抽取:

- 取 100 名简单随机抽样的男性。
- 独立地,取 100 名简单随机抽样的女性。

这是一个较好的设计,因为人为偏差被排除了。

11. 这种设计的细节在第 22 章里讨论;也可参见注 12。我们建议分层必须以使成本合理的方式抽取样本,但是在大多数民意测验中分层几乎不缩减

方差。取一个虚构的例子，假设一个国家由两个地区组成。东部与西部，在东部，60%的选民是民主党；在西部，仅有40%是民主党。东西部在人数上相等，因此民主党的总百分数是50%。现在，两个调查机构抽样去估计民主党的总百分数。第一个机构采用了容量为n的简单随机样本，标准误差是 $50\%/\sqrt{n}$，第二个机构分层，在东部取容量为n/2的简单随机样本。并在西部独立地取容量为n/2的简单随机样本。标准误差是 $\sqrt{0.4\times0.6}\times100\%/\sqrt{n}$。因为 $\sqrt{0.4\times0.6}\approx0.49$，方差的缩减是最低限度的。此外，在这个人为的例子中，地区之间的差异比实际选举中所看到的差异大多了。因此在预测实际选举中分层法的好处甚至更小。对比之下，在抽取经济单位（如公司或企业）时，分层法能真正地有助于减小方差。

12. 盖洛普民意测验采用了随机启动表抽样的变式。在头3个阶段，概率与规模大小成比例；实行中，每个单元以相当于它的大小的倍数出现在表中。在四个地理区域中的每一个内，存在一个农村地区的阶层，对于它的处理多少不同于城市地区。

13. 盖洛普公司指出，“在家庭内选择的这种方法已经经验地被发展而产生按男性和女性各自的年龄分布，而且该分布和总体年龄分布有可比性。”

14. 严格地讲，对于盖洛普民意测验，有可能仅对家庭而不是对个体去计算抽样概率——因为在家庭内选择个体中所采用规则的缘故。不回答是另一种复杂性。我们感谢Ben King(ETS)关于这个问题的有益讨论。

15. Paul Perry,“A Comparison of the voting preferences of likely voters and likely nonvoters,”Public Opinion Quartelg vol. 37(1973)pp. 99—109. 谁参加了选举是一件公开的有案可查之事；当然，他们如何选举则不是这回事。

16. 在头两个阶段，选择的概率与规模大小成比例，大小的测度为电话薄白页中列出的电话数目。这样做是讲得通的，尽管目标总体是由列出的或未列出的全体住宅电话所组成。

17. Kenneth Stampp,California-Berkeley大学历史系荣誉教授。这是一个WPA项目，实验对象必定已经七十多岁了！

18. R. W. Fogel and S. L. Engerman, Time on the Cross:The Economics of American Negro Slavery(New York:W. W. Norton & Company,1989,p.

39;Evidence and Mettods,Little,Brown & Co,1974,p. 37)。细微的评论参见 Richard Sutch,"The treatment received by American slaies,"Explorations in Economic history vol. 12(1975)pp. 335—438。

19. L. Belmont and F. Marolla,"Birth-order,family-sige,and intelligence," Science vol. 182(1973)pp. 1096—1101,平均来说,即使在控制了家庭背景以后,智力随着出生次序和家庭大小而减少。也可参见 R. B. Zajonc,"Family configuration and intelligence,"Science vol. 192(1976)pp. 227—36。
20. L. L. Bairds,The Graduates (Princeton:ETS,1973)。
21. 由医学研究委员会(The Medical Research Council)的 A. L. Cochrane 的讨论,The Application of Scientific Mettods to Industrial and Service Medicine (London:His Majesty's Stationery Office,1951,pp. 36—39)。
22. A. C. Nielsen,1987Annual Report on Television。
23. 根据 Parten 书(注 2)中的一个例子。
24. 摘自 Time on the Cross(注 18)。Anne Arundel 是 Maryland(马里兰)第二个领主Cecil Calvert 的妻子。当时两个主要的奴隶拍卖所在 Annapolis (Arundel County)和 Charleston(South Carolina)。我们感谢(Sharon Tucker 有关 Maryland 历史方面的帮助。
25. E. K. Strong,Japanese in California(Stanford University Press,1933)。
26. San Francisco Chronicle (旧金山纪事)。Dec. 10,1987;由Stephen Peroutka 写给 New England Journal of Medicine vol. 317(1987)pp. 1542—43 的信。
27. 这个例子是由 D. Kahneman and A. Tversky 提出的,"Judgment under uncertainty:heuristics and bias,"Science vol. 185(1974)pp. 1124—31。

第 20 章. 抽样中的机会误差

1. 数据来自健康检查调查的 Cycle I,有关美国年龄为 18—79 岁人的一个概率样本。该研究在 1960—1961 年进行。样本容量为 6672,其中 3091 名为男性。
2. 关于全美国的数据可从Statistical Abstract 1988 得到:表 46 给出了婚姻状况;表 203 给出了年龄 25 岁以上人的教育水平;表 669—704,给出了家庭收入;以及表 489,关于个人税申报的数据。
3. Sir Arthur Conan Doyle, The Sign of The Four (J. B. Lippincott,1899; Ballantine Books,1974,p. 91)。Holmes 把这种思想归功于 Winwood Reade.
4. 380 US 202(1965)。在后来的案件中,法庭认为陪审团必须是社会的代表:

随机抽样给出了社会百分数组成的容许偏差。最主要的案件是 Castaneda 430 US 482(1977),尤其关于 496 的注 17;也可参见 Avery 345US559 (1953),在那里最高法院法官 Frankfurter 认为"法院的头脑,不仅仅是其眼睛都应该是盲目的(指不偏不倚—译者)而把这种情况归于纯粹的偶然性。"在 Batson476 US79(1986)最高法院将权力缩小为断然的拒绝。

"标准差分析"现在不仅使用于选择陪审团的情况,而且用于职业歧视诉讼,甚至用于反垄断事件。McCleskey (107 S Ct 1756,1987)指出,不管怎么样,在极刑案件中法庭更不情愿接受鉴别的统计依据。关于不同观点的统计依据的讨论,参见

B. Black,"Evolving legal standards for the admissibility of scientific evidence,"Science vol. 239(1988)pp. 1508—12;vol. 241(1988)pp. 1413—14。

S. Fienberg,ed. The Evolving Role of Statistical Assessments as Evidence in the Courts (Springer-Verlag,1989)。

M. Finkelstein,"The application of statistical decision theory to the jury discrimination cases,"Harvard Law Review vol. 338(1966) pp. 353—56。

M. Finkelstein and B. Levin,Statistics for Lawgers(Springer-Verlag,1990)。

P. Meier,J. Sacks,and S. L. Zabell,"What happened in Hazelwood: statistics,employment discrimination,and the 80% rule,"American Bar Foundation Research Journal (1984)pp. 139—86。

D. W. Peterson,ed. "Statistical inference in litigation,"Law and Contemporary Problems,vol. 46. no. 4(1983)。

D. L. Rubinfeld,"Econometrics in the courtroom,"Columbia Law Review vol. 85(1985)pp. 1048—97。

第 21 章 百分数的准确性

1. Sir Arthur Conan Dogle,A Study in Scarlet (J. B. Lippincott,1893;Ballantine Books,1975,p. 136)。
2. 有关全美国的数据,见 Statistical Abstract,1988,表 13 与 236。
3. 对于全美国按户主的受教育水平的家庭收入数据在 Statistical Abstract,1988,表 703 中报导。

4. Censas of Manufactures, 1982, Subject Series, General Summary, Part 2: Industry Statistics by Employment Size of Establishement. 习题 5 和 6 多少与这些数据相匹配。
5. 本书采取严格的频率论的观点。其它的观点引证在第 13 章的注 1 中。许多同僚感到在这一节中我们闭上了眼睛穿越智力的布雷区。我们希望他们在关于结果的判断中是宽厚的。
6. 该图的提出是由 Juan Ludlow, CIMASS/UNAM, Mexico。
7. 对于方便样本也常常计算适用于简单随机抽样的标准误差。在某些场合，结果可能是有用的。例如，目的可能是去证明争议中的抽样方法相当地不同于简单随机抽样(如在 P. 411 上的习题 11)。进而，在许多场合关于简单随机样本的标准误差被认为提供了机会误差的较低界限。
8. Statistical Abstract, 1988, Table1227。
9. D. Ravitch and C. E. Finn, Jr, What Do Our 17-Year-Olds Know? (Harper & Row, 1987)。样本是慎重指定的。
10. 有关全国性数据，参见 Statistical Abstract, 1988, Table 879。
11. Keno 的规则在第 17 章的注 1 中解释。单个数的机会是 20/80，因为有 20 次抽取。两个数的机会是(20×19)/(80×79)。

第 22 章　估量就业与失业

1. 我们感谢 Paul Bettin, Margaret Hill, Charles Jones, Donna Kostanish, 和人口局 Jay Waite 的帮助。应当提及 Earl Gerson(已退休)和 Daniel Levine (CNSTAT), 以及 Morton Boisen 和 Julius Shiskin(已亡故)。

   人口普查局负责抽样设计，采集和产生数据，同样地负责估计和其标准误差的计算。劳工局统计进行季节调整，并负责结果的发表和经济解释。关于现场人口调查的若干有用的参考文献：

   Bureau of the Census, The Current Population Survey—A Report on Methodology, Technical Paper No. 7。

   Bureau of the Census, The Current Populatcon Survey: Design and Methodology, Technical Paper No. 40(January, 1978)。

   Bureau of Labor Statistics, Handbook of Methods, Bulletin No. 1171 (1971)。

   Employment and Earnings, Vol. 36, No. 7(July 1989)。

   President's Committee to Appraise Employment and Unemployment

Statistics, Measuring Employment and Unemployment (Washington, D.C, 1962)。

M. Thompson and G. Shapiro, "The current population survey: an overview," Annalsof Economics and Social Measurement vol.2(1973)。

2. 在东北部，有少许例外。

3. 有 23 个特殊的循环 PSU。例如，Montana 有一组 4 个循环县，McCone 是第 1 个；1987 年四月 Petroleum 代替 McCone；1988 年八月 Carter 代替 Petroleum；而 1993 年四月 Danids 代替 Carter。

4. 存在一些例外的大的 USU 它们被不同地处理，在一个 PSU 内，实际通过分层法编组 USU 的名单以缩减方差。抽样因子在八十年代后期大约是正确的，但在 1990 年人口普查之后抽样重新设计时将发生变化。

5. 在 11 个最大的州中每月估计的精确度相等的，在余下 40 个州的年平均的精确度相等。(为此目的，Columbia 特区算作一个州。)这里，"精确度"意指估计的失业人数的变异系数。

6. 监狱，精神病院中的同住者，和武装部队成员都被排斥在"民间的、非公共机构的人口"之外。

7. 总劳动力等于民间劳动力加上军人。

8. 由劳工局使用的方法更复杂了一点。因为他们还用其他人口统计变量来交叉核实。此外，他们利用人口普查数据作出调整的修正样本 PSU 与国家之间已知的人口统计差异；并且他们利用上月的样本信息对现在的估计作出另一调整。获取权的方法有时称作"比估计"。

9. 用于七十年代的方法包括了首先将估计线性化，并用半样本方法计算某些构造区块的方差。由 R. S. Woodruff. "A simple method for approximating the variance of a complicated estimate, Journal of the American Statistical Association" vol. 66(1971) pp. 411—14 所概述。一个完整的描述在 Technical Paper No. 40 中可得到。

10. 分层法缩减了标准误差，如同使用比估计一样。但整群抽样则提高标准误差。

11. Statistical Abstract, 1988, Tables 397 与 418。

12. R. E. Fay, J. S. Passel, J. G. Robinson, and C. D. Cowan, The Coverage of the Population in the 1980 Census, Bureau of the Census, 1988。

13. 证据指出了人口普查局可以很容易地找出谁有全日的工作，以及谁在劳

动力之外。问题是关于第三个组，被分类为非全日工作的工人，或有工作但不上班，或失业的边缘工人。例如，从1987年最后一个季度的再访问计划获得结果可以列表如下：

原先访问时的劳动力状况

再访问时的劳动力状况	农业	在非农业产业中就业			失业	不属劳动力	总和
		全日	部分时间	不属工作之列			
农　业	117	1	2	0	0	2	122
非农业							
全日	0	2 967	22	2	3	3	2 997
部分时间	1	45	1 187	5	4	28	1 270
不工作	0	2	0	137	2	4	145
失业	0	0	2	2	226	21	251
不属劳动力	0	0	2	1	7	2 716	2 726
总　和	118	3 015	1 215	147	242	2 774	7 511

注：在调整之后，加权之前；75%的样本。

来源. Bureau of the Census, Statistical Methods Division。

于是，7 511 人被再访问；在原来的访问中 3015 名被归类为在非农业的行业中全日制工作，但是在再访问中 2 997 名以那种方式归类——大概是正确的。减少了10%的0.6。另一方面，部分时间工人数上升了约4.5%，而失业人数上升了3.7%。失业的总数——基于原来的访问——估计为大约7 000 000。因为7 000 000的3.7%=250 000，估计中的偏差总计有数十万人。在这些数据中失业人数小，因此计算仅仅是阐述想法。

也可参见 K. W. Clarkson and R. F. Meiners, "Institutional changes, reported unemployment, and induced institutional changes," Supplement to Journal of Monetary Economics (1979)增刊。

14. 理论上，整群抽样和比估计可以引起小偏差。但是，实际上，对于相当大的样本来自于这两个来源的偏差是可忽略的。

15. 基于 Hyman 的书(第19章的注2)中的一个例子。

16. A. Jensen, "Environment, heredity and intelligence," Harvard Educational Review (1969, p. 20)。引文稍有剪辑。

第 23 章. 平均数的精度

1. 抽取可以是有放回，或无放回的。在第二种情况，抽取次数——和留在盒中的票数——两者必须大；计算 SE 时可能需要较正因子(第 20 章第 3 节)。参见 T. Höglund,"Sampling fron a finite population;a remainder term estimate,"Scandinavian Journal of Statistics vol. 5(1978)pp. 69—71。
2. Statistical Abstract,1988,Table 236 给出了全美国按年龄和性别的登记人数。
3. Statistical Abstract,1988,Tables 1221—1224 给出有关住房存货的信息。更详细的资料，参见 Current Housing Reports。
4. Statistical Abstract,1988,Table 878 给出有关全美国的数据。
5. Statistical Abstract,1988,Table 1 005 给出有关全美国的数据。另外的信息可以从住宅能源消耗调查(Residential Energy Consumption Survey)的运费调查和从联邦高速公路局私人运输调查(Federal Highway Administration Personal Transportation Survey)中得到。
6. 研究在 1976 年进行。中学课程学年的及格分调查的先前阶段在 Martin Trow,ed. Teachers and Studeuts (NcGraw-Hill,1975)中讨论。采用了分层抽样。在 1985 年，大约有 3 300 所学校，平均入学人数仍大约为 3 700 人：Statistical Abstract,1988,Table 235，也可参见 Digest of Educational Statistics. 然而，通过不同的来源，定义不能准确地比较。
7. 参见注 6，以及 Trow 书的 pp. 6—7。抽样是分层的。

第七部分. 机会模型

第 24 章. 测量误差模型

1. W. J. Youden, Experimentation and Measurement (Washington: National Science Teachers Association,1962).
2. 这样的装置由 Toledo Scale 制造；它采用了 4 个荷载元件，当汽车穿过称重桥时速度最多可到每小时 6 哩。
3. 我们感谢 Sam Cohen 与 Doug Hale (Energy Information Administration)，他们提供了数据。
4. 误差盒有点儿复杂化：95%的票遵循平均数为 0 和 SD 为 4 微克的正态曲

线；其余5%遵循平均数为0和SD为25微克的正态曲线。必需有两条正态曲线，一条是关于中间部分的而一条是关于异常数值的。

5. 根均方(root-mean-square)也许更好一些，因为盒子的平均数假定为0。

6. 重复测量之间的相依性常常由观察者偏差而引起。作测量的人潜意识地希望第二次测量接近于第一次。度量局做了熬费苦心的预防以消除这类偏差。例如，NB10的值通过比较不同砝码组的总质量而得到。这些组按照由度量局发明的设计而变化。真正作测量的人不知道这些组如何与另外的组相关联，因此他对于测量"应当"读作什么没有形成任何观点。

   以更为技术的标准，关于NB10数据的经验分布是非对称的和长尾的(第6章中的图2)，但是，这个分布自身的100重卷积非常接近于正态；与正态的最小偏差通过阶为1/n的Edgeworth展开非常好地得到描述。

7. W. J. Youden,"Enduring values"Technometrics vol,14(1972)pp. 1—11，也可参见M. Henrion and B. Fischhoff,"Assessing uncertainty in physical constants,"American Journal of Physics vol,54(1986)pp. 791—97.

8. 由Michelson,Pease,and Pearson 1929—33年在Irvine Ranch测得。在习题中结果有点四舍五入。他们的关于光速的平均值，转换于每秒英里的话，大约是186 270。测量在若干组中取得，并且有迹象表明误差SD随不同的组而发生变化。

   实质上，现在光速这样定义："1983年the General Conference on Weights and Measures正式地重新确定，光在真空中1/299792,458秒传播的距离就是1米。"参见E. M. Purcell,Electricity and Magnetism,2nd ed. (McGraw-Hill,1985,Appendix E).

9. 引文来自R. D. Tuddenham and M. M. Snyder,Physical Growth of California Boys and Girls from Birth to Eighteen Years (Berkeley:University of California Press,1954,p. 191). 这里有些剪辑。正如作者所继续讲述的，

   > 由于事后的聪明，在研究的后来几年内我们认为理论"真值"的更精确估计将不是第一次的测量记录，甚至也不是'最有代表性的'测量记录，而是简单的一组测量记录的[平均数]。

第25章. 遗传学中的机会模型

1. 我们感谢来自California-Berkeley大学遗传学系的Everett Dempster与Michael Freeling的内行忠告(我们采用了其中的一些)。纽约州农业试验站的G. A. Marx与D. K. Ourecky也是极其有帮助的。最后，我们感谢明

尼苏达大学的 Ann Lane. 两本综合的参考书是：

I. M. Lerner, Heredity, Evolution, and Society (W. H. Freeman, 1968);

E. Rosenberg, Cell and Molecular Biology (Holt, Rinehart and Winston, 1971).

两篇中等的参考书是：

W. T. Keeton and C. H. McFadden, Elements of Biological Science, 3rd ed. (W. W. Norton & Company, 1983);

M. W. Srtickberger, Genetics, 3rd ed. (Macmillan, 1985).

有关分子遗传学的教科书包括：

J. D. Watson, N. H. Hopkings, J. W. Roberts, J. A. Steitz, and A. M. Weiner, Molecular Biology of the Gene, 4th ed (Benjanain Cummings, 1987);

B. Alberts, D. Bray, J. Lewis, M. Raff, K. , Roberts, and J. D. Watson, Molecular Biology of the Cell, 2nd ed. (New York: Garland Publishing, 1989).

肿瘤生物学提供了遗传学中迷人的见识。概述这方面的两本书是：

J. Cairns, Cancer: Science and Society (W. H. Freeman, 1978);

L. M. Franks and N. Teich, eds, Introduction to the Cellular and Molecular Bidogy of Cancer (Oxford University Press, 1986).

2. 严格地说，这仅涉及种子的一部分，子叶或第一批叶子。

3. 术语“基因”稍有点模棱两可，但是它指为蛋白质或主要蛋白质链编码的一个部位的脱氧核糖核酸。孟德尔指控制表型的“本质”。

4. 精子为花粉所带有，卵则在胚珠中，技术上说，这些是细胞核，不是细胞。

5. 基因在染色体上的位置由 Lamprecht (Agric. Hortique Genetica, 1961) 作出。在英国有一个更新近的讨论由 Blixt (同样的期刊，1972) 进行。

6. 《植物杂交实验》(Oliver & Boyd, 1965, p. 53)。该书重印了 Mendels 原来的文章，和基于 Annals of Science vol. 1 (1936) pp. 115—37 上一篇文章的 Fisher 的若干评注。有些遗传学家对 Fisher 的推理感到非常不满；例如，参见 F. Weiling, “What about R. A. Fisher's statement of the ‘too good’ data of J. G. Mendels' Pisum paper?” Journal of Heredity vol. 77 (1986) pp. 281—83。然而，总体来说，Fishers 的论证似乎是有说服力的。

7. 这个实验使用了 5 个特征，而不仅仅是这里讨论的一种。一次试验被重复，

因为Mendel 认为拟合得不好。他在每次试验中使用了100 株植物，共用形成了课文中所提到的600 株。

8. "On the Correlation between relatives on the assumption of Mendelisan inheritance,"Transactions of the Royal Society of Edinburgh vol. 52. pp. 399—433。

9. Biometrika(1903)因子1.08 多少对身高中的性别差异进行了调整。等式是从一篇文章中的一个等式中润色而得。

10. 在这里选择不能起作用；研究总体内1 078 个确定的家庭单元组成。

11. 为从等式(4)得到等式(5)，取给定父亲身高时的条件期望。于是，等式(5)是一个经验事实，附加的模型的一个推论也一样——假定随机配对。

12. 该讨论忽略了类似变种和遗传学家称作交换体的复杂现象。

13. 此习题是从 Strickberger 改编而得。

14. Rasmusson,Hereditas vol. 20(1935)。该问题也来自 Strickberger。

第八部分. 显著性检验

第26 章. 显著性检验

1. 引自 Jerome Frank,Courts on Trial (Princeton Univtersity Press,1949)中的"Of experience"。

2. 微模拟模型多多少少同描述的一样，且被用来评估税法中的变化。典型地，他们不考虑行为的反应。当然，对话是虚构的。

3. Statistical Abstract,1988,Tablc 489 给出了关于个人税申报表的概括性数据。对于1985 年的申报表，平均交税大约为3 000 美元，SD 大约为15 000 美元：基于差异的估计值被简单地加以说明。对于如此的长尾，可以考虑稳健性方法；参见P. Huber,Robust Statistics (John Wiley & Sons,1981)。

4. 其他的读物(以困难程度为序)：

   M. J. Moroney,Facts from Figures,3rd ed。(Penguin Books,1956).

   J. L. Hodges,Jr. and E. Lehmann,Basic Concepts of Probability and Statistics,2nd ed。(Holden-Day,1970)。

   L. Breiman,Statistics with a View towards Applications (Houghton Mifflin,1973)。

P. Bickel and K. Doksum, Mathematical Statistics (Holden-Day, 1977)。

E. Lehmann, Theory of Point Estimation (John Wiley & Sons, 1983)。

E. Lehmann, Testing Statistical Hypotheses, 2nd ed (John Wiley & Sons, 1986)。

5. 我们使用超感官知觉模糊地去包括由于思想作用而影响客观事物的能力和超人的洞察力。实验被描述在 C. Tart Learning to Use Extrasensory Perception (University of Chicago Press, 1976). 事实上，一个实验对象不完成所有的运转。还有结果来自于"十个选择训练者。"

6. 出于课文中给出的理由，我们考虑将备择假设公式化为从一个 1 的比例大于 1/4 的 0-1 盒子中的抽取是不合适的。更合理的可能是说明正确猜测次数随机地大于从盒子 | 1 | 0 | 0 | 0 | 作 7 500 抽取所得之和。

7. 一个这样的实验由 California-Berkeley 大学心理学系的 W. Meredith 教授所指导实施。

8. A. N. Doob et al, "Effect of initial selling price on subsequent sales," Journal of Personality and Social Psychology vol. 11 (1969) pp. 345-50



9. 有关 Student 的轶事在 W. J. Youden, Experimentation and Measurement (Washington; 1962) 中报导.

t-检验是最为流行的统计技巧之一，我们实在很抱歉我们不得不在既枯燥乏味又部分假设的环境里介绍它（课文中的故事直到他们作 t-检验之前都是真实的；实际中他们并没有作 t-检验。）我们没有偶然碰到任何例

子，它们同时是真实的，有趣的，且似乎又是可能的。

对于一个大样本，仅仅需要 z-检验。在小样本的情况，我们的困难如下，t-检验用于计算尾部面积。但是对于正态性的某些偏离使计算形成一个大因子的偏差。于是，为了信赖 t-检验，似乎有必要对误差的分布知道相当多，而没有关于它们的散布的清楚概念——技巧将不使用于那种情况。

对于大样本，关于正态性的偏离没有如此要紧；这是一位统计学家用"t-检验的稳健性"所指的事情。以我们的术语，稳健性这个概念用于 z-检验。也可参见下面的注 11。

从独立性的小小偏离可能对 z-检验与 t-检验两者均有大的影响。

10. 基本上，这只是一种惯例：因子$\sqrt{n/(n-1)}$可以被吸收进由 student 曲线导出的乘数中去。许多计算机产生 SD^+ 而不是 SD：pp. 83。

11. 曲线的公式是

$$y=c\left(1+\frac{t^2}{d}\right)^{-(d+1)/2}$$

$$常数\ c=100\%\frac{\Gamma\left(\frac{d+1}{2}\right)}{\sqrt{\pi d}\,\Gamma\left(\frac{d}{2}\right)}$$

d= 自由度

Γ=Eulers 的伽码函数

Gossett 的方法由 R. A. Fishers 置于严密的数学基础之上，Fisher 也证明了即使当误差不完全遵循正态曲线时该方法可以给出好的近似；这一小样本性质也叫作稳健性(注 9)。两篇参考文献：

H. D. Posten，"The robustness of the one-sample t-test over the Pearson system，"Journal of Statistical Computation and Simulation vol. 9(1979)pp. 133-49；

E. Lehmann and W. Y. Loh，"Pointwise vs. uniform robustness of two statistical procedures，"Scandinavian Journal of Statistics vol. 17(1990)

12. 如果盒中的票遵循正态曲线，那末抽取所得之和的概率直方图也是如此——即使只有少量抽取。技术上，正态曲线自身的卷积给出另一正态曲线。如果盒中的票有一个已知的分布，它不是正态的，利用卷积，统计学家可以作出抽得之和或平均数的概率直方图。

13 A. W. Astin，K. C. Green，W. S. Kom，and M. Schalit，The American

Freshman: National Norms for Fall 1986 (American Council on Education, UCLA, 1987).

14. 参考文献：

Hans Zeisel, "Dr. Spock and the case of the vanishing women jurors," University of Chicago Law Review vol. 37(1969)pp. 1—18。

——, "Race bias in the administration of the death penalty: the Florida experience。" Harvard Law Revivw vol. 95(1981)pp.456—68。

在 Zeisel 发表了 1969 年的文章之后，由法官 Ford 所选的下一组陪审团有 24% 的女性。

15. S. C. Truelove, "治疗试验"在 L. J. Witts, ed Medical Surveys and Clinical Trials (Oxford University Press, 1959)。

使随机化相形之下，暗淡无色的讨论在 T. C. Chalmers, P. Celano, H. S. Sacks, and H. Smith, Jr, "Bias in treatment assignmeat in controlled cli nical trials," New England Journal of Medicine vol. 309 (1983) pp. 1358—61。

16. Statistical Abstract, 1988, Tables 202 与 382 给出全国性数据。市场调查是虚构的。

17. 这些数据是由纽约的公共卫生部首先得到的。我们从 Northwestern University (西北大学)的统计学教授 Sandy Zabell 那儿获得它们。参考文献是 A. J. gzenman and S. L. Zabell, "Babies and the blackout: the genesis of a misconception," Social Science Research vol. 10(1981)pp. 282—99.

明显地，在熄灯事件后 9 个月，于 8 月 8 日星期一，和 8 月 9 日星期二，New York Times (纽约时报)派遣了一名记者去附近的几家医院。医院报告了他们的产科病房比往常忙碌——大概是因为周末工作清淡而星期一与二忙碌的一般模式。这些"发现"刊登在 1966 年 8 月 10 日星期三的头版，文章的标题是"Births Up 9 Months After the Blackout。"这看来是婴儿激增神话的起因。

18. 参考文献：

Statistical Abstract, 1988, section 5。

Criminal Victimization in the United States (U. S. Bureau of Justice Statistics, annual)。

Crime in the United States (FBI, annual)。

全国犯罪调查中的样本是精心设计的，但是在精确性方面约等价于给定容量的简单随机样本。数字已经四舍五入化了，但比较似乎是有效的。

19. S. M. Stigler, "Eight centuries of sampling inspection: the Trial of the Pyx," Journal of the American Statistical Association vol. 72(1977)pp. 493—500。我们的描述有点因袭时尚。

20. M. Rosenzweig, E. L. Bennett, and M. C. Diamond, "Brain changes in response to experience" Scientific American vol. 210(1964)PP. 22—29。事实上，实验使用的不是一对而是三重的，随机地指派去丰富了的，标准的，和剥夺了的环境组。

第 27 章 再论平均数的检验

1. The Third National Mathematics Assessment: Results, Trends and Issues (Princeton: ETS/NAEP, 1983)。平均得分如同所报告的；样本较大，且为精心设计的；SD 是近似的。在 1986 年有一点儿逆转. 参见 The Mathematics Report Card(Princeton: ETS/NAEP, 1988)。

2. 有关全国性数据，参见 A. W. Astin, K. C. Green, W. S. Korn, and M. Schalit, The American Freshman: National Norms for Fall 1986(American Council on Education, UCLA, 1987)。

3. 除了在课文中提到的检验之外，三种其它的检验广泛地使用于这类问题：

(i)在原假设下，两个盒子里的 1 的百分数相同，并且可以用混合两个样本的办法给予估计：

$$\frac{107+132}{200+300}=\frac{239}{500}\approx 48\%。$$

在这基础上，两个盒子的共同 SD 估计为

$$\sqrt{0.48\times 0.52}\approx 0.50$$

这个合在一起的 SD 可以用来计算检验统计量分母中的 SE。(然而，这个合在一起的 SD 不能用于其它目的，譬如作置信区间于差异附近。)

(ii)Fisher 关于两样本中 1 的个数总和的"精确"检验条件，检验统计量是在(假设)第一个样本中 1 的个数，原分布是超几何分布。

(iii)χ^2-统计量可以从 2×2 表中计算而得，并涉及到一个 χ_1^2-分布。

由两样本中 1 的总数的条件下，检验统计量(i)是第一组样本中 1 的个数的单调函数；因此检验(i)和(ii)是等价的。实际上，超几何分布已经由正态作为近似，对相当大的样本近似是不错的。同样地，(iii)等价于(i)和

(ii)；事实上，$\chi^2 \equiv z_1^2$，其中 z_1^2 是以合在一起的SD所计算的z-统计量。

如果z是从分开的SD所计算的z-统计量，如在课文中那样，那末 $Z^2 > Z_1^2$。然而，就象下面将证明的那样，在两样本中1的总数的条件下，Z是第一组样本中1的个数的单调函数，因此我们的检验等价于检验(i)，(ii)和(iii)。

某些记号将是有用的。设 ξ 为一枚p-硬币抛n次中正面（头像）的次数，$\hat{p}=\xi/n$。设 ζ 为一枚q-硬币抛m次中正面（头像）的次数，$\hat{q}=\xi/m$。抛两枚硬币是独立进行的。原假设是 $p=q$；备择假设是 $p \neq q$。我们规定 $s=\xi+\zeta$。设 $r=s/(m+n)$。这个r将被认为是常数。

对于(i)，方差计算为

$$\nu = r(1-r)\left(\frac{1}{n}+\frac{1}{m}\right)$$

而检验统计量是

$$Z_1=(\hat{p}-\hat{q})/\sqrt{\nu}$$

对于我们的检验统计量，方差计算为

$$w=\frac{\hat{p}(1-\hat{p})}{n}+\frac{\hat{q}(1-\hat{q})}{m}$$

且检验统计量是

$$Z=(\hat{p}-\hat{q})/\sqrt{w}\ 。$$

我们把m，n，和 $r=s/(m+n)$ 看成为是固定的；策略上，我们已假定 $0<\hat{p},\hat{q},r<1$。引进新的变量 $x=\hat{p}-r$ 将是方便的，从而 $\hat{q}=r-\frac{n}{m}x$。显然，x仅取有限许多个值；然而，在以后将x认为从它的最小值到它的最大值遍历整个区间是于事方便的。总结变量的变化，

$$\hat{p}=r+x,\quad \hat{q}=r-\frac{n}{m}x。$$

于是

$$\hat{p}(1-\hat{p})=r(1-r)+(1-2r)x-x^2$$

$$\hat{q}(1-\hat{q})=r(1-r)-\frac{n}{m}(1-2r)x-\frac{n^2}{m^2}x^2$$

$$w=\nu+bx-cx^2$$

其中

$$b=\left(\frac{1}{n}-\frac{n}{m^2}\right)(1-2r),\quad c=\frac{1}{n}+\frac{n^2}{m^3}$$

最后，

$$Z=\frac{m+n}{m}x\Big/\sqrt{\nu+bx-cx^2}。$$

当然，w，ν，与 c 全为正数；b 可能为正或负。作为 x 的函数 z 的单调性遵循如下引理。

引理，设 ν,c>0 和设 b 为实数，限制 x 的区间满足

$\nu+bx-cx^2>0$，设

$$f(x)=x/\sqrt{\upsilon+bx-cx^2}。$$

那末 f(x)关于 x 为单调递增。

证明，如果 b≤0，这是平凡的；因此设 b>0，则

$$\frac{df}{dx}=\frac{2\nu+bx}{2(\nu+bx-cx^2)^{3/2}}>0$$

4. Literacy among youths 12—17 years,"Vital and Health Statistics Series 11,No 131(Washington,D.C.,1973)。样本设计类似于现场人口调查，且调研人员利用半样本法估计标准误差。习题中所指示的容量的简单随机抽样将具有大约等同于实际状况的标准误差。
5. "Intellectual development of children by demographic and socioeconomic factors,"Vital and Health Statistics Series 11,NO. 110(Washington,D.C. 1971)。有关标准误差的讨论，参见注 4。小孩的考试得分与父母的教育之间的相关系数是 0.5，当父母的收入维持常数时下降到 0.3。"大城市"是指三百万及以上的人口。事实上，在具有人口范围从 1 至 3 百万的城市中的儿童干得最好，平均大约 28 分。
6. Statistical Abstract, 1988, Table 180. U. S. National Institute on Drug Abuse。National Household Survey on Drug Abuse。百分数和样本容量有稍微的变动。抽样是精心设计的。

   也可参看 National Trends in Drug Use and Related Factors among American High School Students and Young Adults,1975-1986。(Washington,D.C. National Institute on Drug Abuse,1987)。在整个时期，可卡因使用的"三十天流行"增加了；使用大麻似乎在二十世纪七十年代后期达到高峰。
7. 参见注 2 中的来源。
8. 有关对维生素 C 的实际试验的参考文献，对于使随机化盲化带有有趣的

间接说明，可参见第 2 章中的注 10。

9. 考虑一个临床试验来比较疗法 A 与 B。假设有 N 个受实验对象，依次标为 $i=1,2,\cdots,N$。设 x_i 为对象 i 的对于疗法 A 的反应；同样地，y_i 是对象 i 关于疗法 B 的反应。对每一个 i，或者 x_i 或者 y_i 可以观测到，但不可能同时观测到两者，设

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i \qquad \bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i$$

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2 \qquad \tau^2=\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2$$

$$\operatorname{cov}(x,y)=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$$

因为 x 与 y 不是同时可观测的，$\operatorname{cov}(x,y)$ 不能用一个基于样本的协方差来估计。其余四个参数可以从样本数据中得到估计。

设 S 是 $\{1,\cdots,N\}$ 的一个随机子集，含有 n 个元素；这一组获得疗法 A；因此对于 $i\in S$ 观察到 x_i。设 T 是 $\{1,\cdots,N\}$ 的一个随机子集，含 m 个元素，与 S 相分离，这一组得到疗法 B，因此对于 $j\in T$ 观察到 y_j。我们用样本均值

$$\overline{X}=\frac{1}{n}\sum_{i\in S}x_i \qquad \overline{Y}=\frac{1}{m}\sum_{j\in S}y_j$$

估计总体均值 $\bar{x}$ 和 $\bar{y}$，由组合数学的计算，

$$\operatorname{Var}\overline{X}=\frac{N-n}{N-1}\frac{\sigma^2}{n} \qquad \operatorname{Var}\overline{Y}=\frac{N-m}{N-1}\frac{\tau^2}{m}$$

$$\operatorname{cov}(\overline{X},\overline{Y})=-\frac{1}{N-1}\operatorname{cov}(x,y)$$

于是

$$\begin{aligned}\operatorname{Var}(\overline{X}-\overline{Y})&=\frac{N-n}{N-1}\frac{\sigma^2}{n}+\frac{N-m}{N-1}\frac{\tau^2}{m}+\frac{2}{N-1}\operatorname{cov}(x,y)\\&=\frac{N}{N-1}\left(\frac{\sigma^2}{n}+\frac{\tau^2}{m}\right)+\frac{1}{N-1}\left[2\operatorname{cov}(x,y)-\sigma^2-\tau^2\right]\\&\leqslant\frac{N}{N-1}\left(\frac{\sigma^2}{n}+\frac{\tau^2}{m}\right)\end{aligned}$$

因为 $\operatorname{cov}(x,y)\leqslant\sigma\tau$ 和 $2\sigma\tau-\sigma^2-\tau^2\leqslant0$. 在课文中提出的“保守估计”是 $\frac{\sigma^2}{n}+\frac{\tau^2}{m}$；其中 σ^2 和 τ^2 将用样本方差来估计。

符号可能有点儿困惑。一般，我们期望 x 和 y 在所有的实验对象范围

是正相关的。如果过多的具有高 x-值的实验对象被指定去疗法组 A，那末太少的具有高 y-值的实验对象留在疗法组 B。因此样本平均数 $\overline{X}$ 和 $\overline{Y}$ 是负相关的。通常，cov(x，y)应当接近它的上限 $\sigma\tau$，因为横穿全体对象的 x 与 y 将是高度相关的。于是"保守"估计应该相当精确，至少对于相当大的样本是如此。有时候，原假设指定 $x\equiv y$。那末 $\sigma=\tau$，而且计算是准确的：参见下面的注 12。当然，如果 N 相对于 m 和 n 而言是大的，于是 $\overline{X}$ 和 $\overline{Y}$ 几乎是独立的；再一次地，"保守估计"将接近于正确。

其他变量的影响可以处理如下。设 η 表示治疗状态。ω 表示影响反应的其它变量的状态。我们假定存在一个函数 f 使得实际对象 i 关于疗法的反应是 $f(i,\eta,\omega)$。令 ρ 表示指派变量：如果 $\rho(i)=A$ 那末实验对象 i 指派去疗法组 A，指派去疗法组 B 是类似的。我们假设 ρ 与 ω 是独立的：给定 ω，在实验对象经指定为 A 的基数(势)为 n 的组 S 与指定为 B 的势为 m 的另一组的所有划分过程中 ρ 的分布律是均匀的。随机化的目的是保证这个假设。于是对每一个 ω，用

$$x_i=f(i,A,\omega) \qquad i\in S$$

$$y_j=f(j,B,\omega) \qquad j\in T$$

可以分开地进行我们的讨论。

一个技术技节问题：有关的中心极限定理适合于无放回抽样(第 23 章注 1)。

模型似乎回到Neyman 关于农业实验的早期工作：参见H. Scheffé，"Models in the analysis of variance，"Annals of Mathematical Statistics vol. 27 (1936)和 J. Neyman，"Sur les applications de la théorie des probabilitiés aux experiences agricoles：Essai des principles，"Roczniki Nauk Rolniczki vol. 10(1923)pp. 1—51，以波兰文。英语翻译为D. Dabrowska and T. Speed，Technical Report No. 235，Statistics Department，University of California，Berkeley，Statistical Scienes vol. 5(1991)pp. 463—80

有关进一步的讨论，参见 J. L. Hodges，Jr，and E. Lehmann，Basic Concepts of Probability and Statistics，2nd ed. (Holden-Day，1970，section 9. 4)。也可参见 J. Robins，"Confidence intervals for causal parameters。Statistics in Medicine vol. 7(1988)pp. 773—85。

10. Amos Tversky and Daniel Kahneman，"Rational choice and the framing of decisions，"Journal of Business vol. 59(1986)no. 4 part 2.pp. S251—78。

11. arbara J. McNeil. Stephen G. Pauker, Harold C. Sox, Jr, and Amos Tversky, "on the elicitation of preferences for alternative therapies," New England Journal of Medicine vol. 306(1982)pp. 1259—62。

12. 我们继续注 3 和注 9 中的讨论。(表 1)总计有 80+87=167 名实验对象。他们中间,40+73=113 名赞成手术治疗;其余 54 名赞成放射治疗。这里,原假定规定 x≡y,因此 $\sigma=\tau$ 且两者均可从数据中计算而得。实际上,在原假设下,赞成手术治疗的医生的百分数是 113/167×100%≈68%,因此

$$\sigma=\tau\approx\sqrt{0.68\times0.32}\approx0.47$$

同样,$\overline{X}$ 与 $\overline{Y}$ 之间的协方差可以精确地计算出来。此项达到上界 $\sigma\tau=\sigma^2$,因为在整个实验对象范围内 x 和 y 之间的相关系数是 1。于是

$$\mathrm{var}(\overline{X}-\overline{Y})=\frac{N}{N-1}\left(\frac{1}{n}+\frac{1}{m}\right)\sigma^2。$$

检验统计量的两种形式(合并的或分开的 SDs)在实质上还是相同的。例如,利用原假设确定模型,在两个统计量之间的值的 r. m. s 差异仅为 0.013。而且,正态近似是十分好的:对任一个统计量,绝对值超过 2 的机会大约是 4.8%,可以和正态尾部概率 4.6%相比。

13. Daniel Kahneman and Amos Tversky. "Choices, values, and frames," American Psychologist vol. 39(1984)pp. 341—50。

14. 事实上,随机化有些复杂。在整个期间预防需要 3 次分开的注射,从而对照组也给予 3 次(安慰剂)注射。含注射材料的小药水瓶 6 个一盒地包装;3 个含有疫苗且有共同的代码数;另 3 个含有安慰剂,具有另外的共同代码数。每个小药水瓶有足够的药水供 10 次注射用。

当第一轮注射的时间到来时,从盒中取出一瓶,10 个小孩从该瓶中获取注射;调研人员对这 10 个小孩记下瓶的代码数;这 10 个小孩从盒中具有相同代码数的另外 2 瓶中得到他们的第二和第三次注射。接下来的 10 个小孩从该盒了中另一组的 3 瓶(具有与第 1 次使用的那瓶不同的代码数)中的 1 瓶得到他们的第一轮注射;对他们记下小瓶的代码数:他们以后的注射来自于该组中余下的 2 瓶。

于是实际上小孩们被划分为若干对,每对二组每组 10 人;对每一对抛一枚硬币;以 50 对 50 的机会,一个整组指派去治疗组,而余下一组则进对照组。课文中的计算是准确的,基于没有两个脊髓质灰炎病例从同一

盒中得到注射这一似乎可能的假设之上。否则，计算必须被修正。这个特殊的试验通常用两样本 z-检验来分析，而不考虑划分(第 1 章的注 2)。我们照着做。

15. D. N. Kershaw and F. Skidmore, The New Jersey Graduated Work Incentive Experiment (Princeton: Mathematica, 1974)。我们乐于感谢数学界的 Rob Hollister 先生的帮助。实际上的实验在其设计中更 复杂化，并且存在显著的样本消耗。也可参见 L. Neuberg, "Distorted transmission," Theory and Society vol. 17(1988)pp. 487—525。

16. D. Ravitch and C. E. Finn, Jr, What Do Our 17-Year-Olds Know? (Harper & Row, 1987, p. 52)。苏联有最高褒奖因素。

17. 参考文献:

K. Gray-Donald, M. S. Kramer, S. Munday et al. "Effect of formula supplementation in the hospital on duration of breast-feeding: a controlled clinical trial," Pediatvics vol. 75(1985), pp. 514—18。

K. Gray-Donald and M. S. Kramer, "Causality inference in observational vs, experimental studies: an empirical comparison," American Journal of Epidemiology vol. 127(1988) pp. 885—92。

在进行对照实验之前，这些调研人员也进行观察研究，在研究中托儿所均按照标准的补给惯例。在托儿所中的补给与以后的哺乳之间存在着一个强负相关，如同在以前的研究中那样。

技术上，对于托儿所的指派不是随机的：当一位母亲出现时，她被分到带有一张可使用的床的托儿所；这事由不得卷入研究的办公室人员来做。合适人选按照在调查书中所指定的客观准则而确定。

调研人员友好地提供了未发表的数据。

18. George Gallup, Jr. The Gellup Poll (Wilmington, Del: Scholarly Resources, 1987)。

19. 参见注 10 中有关来源。

20. The Third National Mathematics Assessment: Results, Trends and Issues (Princeton: ETS/NAEP, 1983)。项目来自于评价，其结果差不多如同报告的那样；计算器组实际上做得较糟糕。但是，从报告中不清楚该项研究是观察地还是实验地进行。

21. P. H. Rossi, R. A. Berk, and K. J. Lenihan, Money, Work and Crime: Ex-

perimental Evidence (San Diego: Academic Press, 1980, 尤其是 Table 5. 1)。研究在1976年进行。我们已经简化了实验设计,但不是在任何本质上的简化。Rossi 等人争辩收入资助缩减了再犯罪,但是其效果被工作周的影响而掩饰了。他们的分析受到 H. Zeisel 严厉的批判,"Disagreement over the evaluation of a controlled experiment," American Journal of Sociology vol. 88(1982)pp. 378—96, with discussion。

22. S. J. Sherman, "On the self-erasing nature of errors of prediction," Journal of Personality and Social Psychology vol. 19(1980)pp. 211—21。

23. 参考文献:

"Multiple Risk Factor Intervention Trial," Journal of the American Medical Association vol. 248(1982)pp. 1465—77。

"Statistical design considerations in the NHLI Multiple Risk Factor Intervention Trial (MRFIT)。" Journal of Chronic Diseases vol. 30 (1972)pp. 261—75)。

随机化包括组的划分,这里不说明原因。平均数是发表了的;SD 由 J. D. Neaton (University of Minnesota, 生物统计学教授)友好地提供。

一个有趣的偶然启示:拟合于 Framingham 数据的逻辑斯蒂回归预测了由于在风险因子上看来有节制的减少(血压减少 3mm, 血清胆固醇下降 5mg/dl, 吸烟减少 13%)从而在死亡率上有大量的缩减。存在着一些忧虑,即治疗组少报了抽烟情况,通过血液化学性质对此作出调整。

也可参见"Mortality rates after 10. 5 years for participants in the Multiple Risk Factor Intervention Trial," Journal of the American Medical Association vol. 263(1990)pp. 1795—1801。

24. Dr. William Epstein, 如在纽约时报 1988 年 9 月 27 日所报导的那样。

第 28 章. χ^2-检验

1. "在一个变量的相关系统情况下,一个已知系统对可能性的偏差是基于这样的一个标准,即它(偏差)能被合理地认为是来自随机抽样。" Phil. Mag. series V vol. 1 pp. 157—75。

2. 如果机会模型是正确的,每一项期望比 1 略小一点;所有项的和期望为 n −1, 其中 n 是项数。

3. 曲线的方程是

$$y=\frac{100\%}{\Gamma(d/2)}(\frac{1}{2})^{\frac{d}{2}}x^{\frac{d}{2}-1}e^{-x/2}$$

d＝自由度

Γ＝Eulers′ Γ-函数

4. 精确分布是在 IBM PC-XT286 上计算而得，采用了 True-BASIC 程序，运行了大约 5 分钟。该程序通过将总计为 60 的以单调顺序排列的所有 6-维数。对每一个 6-维数计算 χ^2统计量和相应的概率(采用多项公式)。这些概率相加给出了答案和图 2 中的概率直方图。计算似乎精确到大约 15 位小数，因为所有概率之和是 $1-10^{-5}$。图 2 中的摆动是真实的。

   许多书推荐 Yates 修正(当绝对差超过 0.5 时，在平方之前从这个差减去 0.5)。对于 1 个自由度，这样做等价于连续性修正(P. 291)且是一件好事。对于大于 1 的自由度，数值计算表明这常常是件坏事。直方图可能非常远地移向左边。数值计算的结果也表明对于平均每格 5 的观测值以及仅少许自由度，χ^2曲线可以信赖到 5%左右，每格平均有 10 个观测值时，它可以相当好地信赖到超过 1%。

5. 在盒中只有两类票时，χ^2统计量等于 z-统计量的平方。因为正态变量的平方是具有 1 个自由度的 χ^2，在这种情况 χ^2检验精确地给出如(双尾)z-检验一样的结果。也可参见第 27 章的注 3。

6. 有关这例题和 P. 484 上习题 9 的数据由 California State Lottery 通过他们的统计顾问 Don Ylvisaker(UCLA)友好地提供。P-值是通过重复的分部积分计算出数值的。通常的正态近似并不起如此好的作用，但是有一个方法令人惊异地给出了精确的结果：D. B. Peizer and J. W. Pratt, "A normal approximation for binomial, F. beta, and other common, related tail probabilities," Jonrnal of the American Statistical Association vol. 63(1968)pp. 1416—83。

7. 在某些情况(例如，每格只有少量观测值)，将数据分组是可取的。

8. UCLA Law Review vol. 20(1973)p. 615。

9. 参见第 25 章的注 6。

10. A. R. Luria, The Working Brain (New York: Basic Books, 1973)。

11. HANES 设计包含了一群样本，因此在数据中存在着某些相依性，χ^2-检验将不会考虑到这。半样本方法可以用来产生原假设下分布。在所有的年龄组里，妇女始终比男人用右手的多。参见 Anthropometrie Reference Data

and Prevalence of Overweight: United States, 1976—1980，数据来自 The National Health Survey, Series 11. No. 238, U. S. Department of Health and Human Services, Washington, D. C. 表 5 中的数字接近于实际数据，并且使得运算易于进行。

12. 未加权的计数来自由 Bureau of the Census（人口普查局）提供的数据磁带。x^2检验没有考虑抽样设计，但是差异是真实的。数字已经简化了一点。

13. UCLA Law Review vol. 20C1973)p. 616。

14. 未加权的计数来自由 Bureau of the Census（人口普查局）提供的数据磁带。χ^2 检验没有考虑抽样设计。在横贯全部年龄组的许多这样的调查中，从未结婚的男人在工作中较少成功。然而，对于妇女，从未结婚者的失业率与结过婚组的大致相等。（下面的表给出了 35—44 岁的数据，来自 1988 年 3 月现场人口调查，不限于 California 州。）

男 人

	结过婚	丧偶、离婚、或分居	从未结婚
就业	7660	1190	856
失业	308	88	67
不在劳动力之列	278	148	175

妇 女

	结过婚	丧偶、离婚、或分居	从未结婚
就业	5879	1798	699
失业	218	120	26
不在劳动力之列	2323	422	163

15. 此习题从 IRRI 提供的数据加以改编。

第 29 章. 显著性检验的更准确的考虑

1. 225 US 391，引文来自 Jerome Frank, Courts on Trial (Princeton University Press, 1949)

2. 当他是 Journal of Experimental Psychology 的编辑时，Arthur Melton 以如下这些话为实际情况辩护：

> 对一篇文章评估的下一步包含了关于在调查结果中所给予信任的判断——即信任在描述的条件下实验的结果是可以重复的。在编辑杂志中，对于同涉及研究中的主要关心有关的结果，当这些结果不管用单尾或双尾检验都（仅仅）在 0.05 水平上为显著的，对此已经存在着强烈地不愿去接受或发表它们的现象。这并没有暗指对于 0.01 水

平或任何其它水平有盲目的崇尚，如某些批评家所暗指的那样。相反，它反映了一种信念，即调研人员对于科学的责任是以这样的方式去显示他的结果使得没有一个讲通理的人将能够通过讲它们是按球反弹方式产生的结果而去怀疑该结果。

存在一个较好的办法保证结果是可重复的，即，坚决要求应该重复重要的实验。引文来自杂志(Vol. 64(1962),pp. 553—57)中的一篇社论。我们在 David Bakan 的一篇文章中发现了它，重印在 Joseph Steger,ed,Readings in Statisties(Holt,Rinehart and Winston,1971)。

3. 历史典故出于 G. A. Barnard, Imperial College of Science and Technology,London,以前的统计学教授；现在已退休。

4. 另外一个好的办法是交叉核实回归：在一半数据上建立模型，然后看看当公式应用于另外一半数据时拟合显示得有多好。重复过程是个关键的想法。在课文中我们没有充分地这样做。

5. 例子是程式化的，但问题是真实的。我们假设了每年每 100 000 分之一的发生率，并且采用了 Poisson 模型。尽管有关于环境污染的担心，自从二十世纪三十年代以来，在美国肝癌率一直稳步地下降。有关讨论和其它的参考文献，参见 D. Freedman and H. Zeisel,"From monse to man:the quantitative assessmaent of cancer risks,"Statistical Science vol. 3(1988)pp. 3—56,with discussion。关于一组的一个争执性例子，参见 S. W. Lagakos,B. S. Wessen, and M. Zelen, "An analysis of contarinated well water and health effects in Woburn. Massachusetts,"Journal of the American Statistical Association vol. 81(1986)pp. 583—614,with discussion。

6. 在其它情况，较难修正数据探索的 P-值，参见 Theo Dijkstra, ed. On Model Uncertainty and Its Statistical Implications,Springer Lecture Notes in Eeonomics and Mathematical Systems,no. 307,1988。对于杂志出版的影响的某些讨论，参见：

   L. J. Chase and R. B. Chase,"A Statistical power analysis of applied psychological research,"Journal of Applied Psychology vol. 61(1976)pp. 234—37。

   K. Dickersin, S. Chan, T. C. Chalmers, H. S. Sacks, and H. R. Smith, Jr,"Publication bias and clinical trials,"Journal of Controlled Clinical Trials vol. 8(1987)pp. 343—53。

A. Tversky and D. Kahneman, "Belief in the law of small numbers, "Psychology Bulletin vol. 2(1971)pp. 105—10。

C. B. Begg and J. A. Berlin, "Publication bias and dissemination of climical research, "Journal of the National Concer Institute vol. 81 (1989)pp. 107—15。

7. "The Lipid Research Climics primary prevention trial results" Journal of the American Medical Association vol. 251(1984)pp. 351—64。基于生命表分析与分块，研究人员引用 $z \approx -1.92$。方案没有说明使用的是单尾还是双尾的检验；它指明"与胆固醇降低剂相关联的值得注意的发病率和死亡率"；并且宣布 1% 的显著性水平"被选来作为显示治疗组之间令人信服的差异的标准。"有一个很强烈的建议，即致命的与非致命的心脏病突发应分别进行分析——在差异是不显著的情况下。参见 Journal of Chronic Diseases vol. 32(1979)pp. 609—31。有一个不太正式的描述见 T. J. Moore, Heart Failure (Rardom Honse, 1989)。近期的实验报导在 H. Buchwald et al 。"Effect of partial ileal bypass surgery on mortality and morbidity from coronary heart disease in patients with hypercholesterolemia, "New England Journal of Medicine vol. 323(1990)pp. 946—955。

8. K. R. Rao, ed, "The Ganzfeld debate, "Journal of Parapsychology vol. 49 no. 1(1985)and vol. 50 no. 4(1986)。分布的离散是重要的，给定的概率是基于卷积程序。

9. 生物测定结果的评价是一件复杂的问题，但是多端点问题是现实的一个。许多化学制品似乎引起肝癌但预防白血病——对于老鼠。参见在注 5 中提及的 Freedman and Zeisel 的论文。

10. T. C. Chalmers, R. S. Koff, and G. F. Grady, "A note on fatality in serum hepatitis, "Journal of Gastroenterology and Hepatology vol. 69(1965)pp. 22—26。

11. "统计显著性"与重要性之间的混淆随着相关系数而变坏。某些研究人员将检验是否 $r=0$，从而采用 P 作为相关联的测度来替代考虑 r 的值。回归系数常常受到同样的处理。然而，正是方差分析以它最敏锐的形式提出问题：有些研究人员将报告 p-值，F-统计量，和除了他们的结果的重要性以外的每一件事。有关一些讨论，参见 P. E. Meehl, "Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of Soft psy-

chology,"Journal of Consulting and Clinical Psychology vol. 46(1978)pp. 806—34。

12. 另一方面，大城市儿童与农村儿童在他们较大年龄时两者之间阅读能力方面可能存在显著的差异。参见 I. S. Kirsch and A. Jungeblut ,Literacy: Profiles of America's Young Adults (Princeton:ETS/NAEP,1986)。
13. 6 分来自于一个粗糙且容易的回归分析，且包括了选择的影响。
14. 这是由 D. T. Campbell,"Reforms as experiments,"American Psychologist vol. 24(1969)pp. 409—29 所作一篇评论(取自上下文)的一个接近的释意。该参考文献是由去世不久的 Merrill Carlsmith 提供的，他是 Stanford University 以前的心理学教授。
15. M. J. Mahoney,"Publication prejudices:an experimental study of confirmatory bias in the peer review system,"Journal of Cognitive Therapy and Research vol. 1(1977)pp. 161—75。实验设计和引语已经简化了一点。
16. Daniel McFadden,"The vevealed preferences of a goverument bureaucracy:empirical evidence,"Bell Journal of Economics vol. 7(1971)pp. 55—72。研究期是 1958—1966 年。模型可能容易受到疑问。此参考文献由 Chris Achen 提供。他是 University of Chicago 的政治学教授。
17. 由 W. Hogan 和 J. Kalt (Harvard)在 1987 年关于违反油价控制的行政意见听取会中提出的证据的释义。
18. 由 C. E. M. Hansel ESP: A Scientific Evaluation (Charles Scribner's Sons,1966,Chapter ll)对实验进行了讨论。数字已被改变以简化运算。实验的关键是解释课文中讨论的谬误。参考文献由 Dr. Charles Yarbrough, Kenwood,California 提供。
19. 关于 Aquarius 本身的随机数产生器似乎没有检验过，但是产生器类似于已被检验过的那些。在 ESP 研究中，没有什么东西是简单的，Tart 将不会同意我们所写中的许多。关于该问题的热烈讨论，参看 Martin Gardner, Science:Good,Bad,and Bogus(Avon Books,1981,Chapters 18 and 31。) 也可参见 Charles Tart et al,"Effecfs of immediate feedback on ESP performance:a second study,"Journal of the American Society for Psychical Research vol. 73(1979)pp. 151—65。
20. 为意释 Keynes(凯因斯)，一个认为他不需要盒子模型的显著性检验者可能恰好有一个天真的盒子。J. M. Keynes,The General Theory of Employ-

ment, Interest, and Money (Harcourt Brace Jovanovich, 1935, pp. 383—84)。

实际经验丰富的人们相信他们自己完全摆脱任何学理的影响，他们通常是某些已死的经济学家的奴隶。

21. Statistical Abstract, 1988, Table 20。

22. 这项研究在第 2 章的第 4 节中讨论过；为参考见：也可参看那章的注 7。

进行算术运算 z≈5 因此 P 相当小。我们可以将 P 理解为描述性统计量。总共有 933 名报考人，他们中间 825 名是男性而 108 名是妇女。如果你认为性别与录取是无关的，比较男生和女生的录取率就象比较任何一组 825 人的录取率与余下的 108 名一组的录取率一样。（毕竟，根据姓、指纹印，诸如此类，存在着许多其它的不相干分法。）将 933 名报考人分为两个组，一组容量 825 人和另一组容量 108 人，有$\binom{933}{825}\approx 7\times 10^{143}$种可能的方式。对每一种分法，计算 z，z-值的这个总体接近于正态分布，因此观察到的 z-值 5 是十分不寻常的，参见 D. Freedman and D. Lane，"A non-stochastic interpretation of reported significance levels，"Journal of Business and Economic Statistics vol. 1(1983)pp. 292—98。

23. Project Follow Through Classroom Evaluation，由在 Menlo Park，California 的 SRI 出版，资深研究者是 Jane Stallings。引文稍有点剪辑。研究是在 1972—1973 年进行。

24. 这假设了对照平均数 60 已知没有误差。事实上，SRI 作了两样本 t-检验。但是，SRI 评分程序应该引进处理和对照分之间的相依性—— 它建立于合在一起的秩上。

25. 这些是真实的数，来自 1976 年。大约一半的 TA 参与期终评分，许多人对其他学期中类似的期终考评了分。

26. F. Mosteller and R. Rourke Sturdy Statistics (Addison-Wesley, 1971, p. 54)。

27. T. A. Ryan, B. L. Joiner, and B. F. Ryan, Minitab Student Handbook (Boston: Duxbury Press, 1976, p. 228)。

28. "Intellectual development of children by demographic and socioeconomic factor，"Vital and Health Statistics Series 11, No. 110(Washington, D. C, 1971)。

29. R. S. Erikson, J. P. Mclver, and G. C. Wright, Jr, "State political culture and public opinion," American Political Science Review vol. 81(1987)pp. 797—813。

分析技巧是关于人口统计学范畴的虚变量的多元回归(例如,低收入,等等);于是关于地区和州的虚拟量被增加。在州虚拟量中添加使调整的 R^2 从 0.0898 增加到 0.0953,但开始的 F(统计量)是 8.35,在分子上有 40 个自由度—— 同时分母上是 55072。作者说在实际条件下州的结果也是显著的;R^2 的建议则不同。作者们承认州虚拟量可以是省略变量的代理量,但反对这种解释。

30. J. L. Gibson, "Political intolerance and Political repression during the McCarthy era。" American Political Science Review Vol. 82 (1988) pp. 511—39。"结果"在道路模型中是系数。作者大概把估计中的随机性看作是由模型产生的。另一方面,模型的充分性可能容易受到疑问。

31. 得到出版者, Harcourt Brace Jovanovich, Inc 的同意而重印。

32. 基于 A. Tversky and D. Kahneman 使用的一个问题. 也见 Steger'S 的书的 P. 298,加注在上面的注 2。

33. 看注 26 中的参考书的 p. 68。

34. F. Arcelus 和 A. H. Meltzer, "The effect of aggregate economic variables on congressional elections," American Political Science Review Vol. 69 (1965)pp. 1232—69,有讨论。

本参考是由 chris Achen 提供的。辩论比习题中指出的较巧妙,并使用了一个回归模型。(模型的有效性当然容易受到疑问)然而,研究者关于假设检验的立场几乎是不讲理的;见 Arcelus 和 Meltzer 对 Goodman 和 Kramer 的评论的回答。

35. Statistical Abstract, 1988, Table 21。

36. Statistical Abstract, 1988, Table 607;包括全日和非全日工作的工人。

37. Statistical Abstract, 1988, Table 237;也见 27 章的注 2。

38. M. S. Kanarek et al., "Asbestos in drinking water and cancer incidence in the San Francisco Bay Area," American Journal of Epidemiology Vol. 112 (1980)pp. 54—72。

对于黑人或妇女来说水中的石棉和肺癌之间没有联系。论文中的数据强烈表明抽烟是一个混淆因素。

39. R. E. Just 和 W. S. Chern, "Tomatoes, technology and oligopsony," Bell Journal of Economics Vol. 11 (1980) pp. 584—602。关于讨论，见 R. Daggett 和 D. Freedman, "Econometrics and law: a case study in the proof of antitrust damages," in L. M. LeCam 和 R. A. Olshen, eds . Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer Vol. 1 pp. 123—72 (Wadsworth, 1985) Just and chern 估计了线性和对数线性两者的需求函数；问题中的统计方法实际指的是线性形式。

40. 见 19 章的注 19。没有兄弟姊妹的孩子是一个例外，得分稍低于两个孩子家庭中第一个出生的孩子的得分。

习 题 解 答

第一部分 实验设计

第二章 观察研究

A 组 第 21 页

1. 不，因为总体也变大了。你应该考虑相对于整个总体容量的死亡数。1985 年总体约为二亿四千万，而在 1960 年它约为一亿八千万：240 分之 2.1 小于 180 分之 1.7，故 1985 年死亡率较低，看起来好象总体实际上变得更健康。
2. NFIP 的对照组有一个完整的家庭背景变化范围。随机化试验中的对照组是取自同意参加的家庭。这些家庭比较富裕，而且小孩容易感染脊髓灰质炎。参见第 1 章第 1 节。
3. NFIP 研究不可能是盲式的；小孩们知道他们是几年级，医生也知道，这可能导致结果偏向有利疫苗（安慰剂达到目的）。
4. 不，因为试验区域是从国内那些对脊髓灰质炎具最大风险的区域中选取的。参见第 1 章第 1 节。
5. 违反了“盲式”规则的人发现了他们服用的是或者不是维生素 C，那些知道他们服用维生素 C 用于予防的人倾向于较少患感冒，那些知道他们服用维生素 C 用于治疗的人倾向于患感冒期较短，这就是安慰剂作用，盲式是重要的。
6. 558/1 096≈51%，1 813/2 789≈65%。在烟酸组里坚持者较少，是随机化——双盲式系统出了差错。（譬如，烟酸可能有不良副作用，引起实验对象停止服用它。）
7. 在试验(i)中，随机化肯定出了点毛病。49.3%与 69.0%间的差表明处理组一开始就比较健康，它将对以后的任何比较产生偏性。差不应归因于处理，因为基线数据指出实验对象在指派到处理组或对照组之前象是那一种状况。（在第 27 章中将有更多这方面的叙述。）
8. (a)是：在处理组里 39 人因乳房癌死亡，对比对照组里有 63 人死亡。此外，在处理组里所有其它原因的死亡人数也较低。

   (b)不，“检查”组与“拒绝”组之间的比较是观察的，即使它是在试验范畴内

进行的；是妇女自己决定是否接受检查，这就象安妥明试验（第 2 节）里执行规定的坚持者。故存在令人担心的混杂因子，诸如收入和教育。

(c)HIP 试验不可能盲式化；妇女们知道她们是否被邀请接受检查。

评论。(i)为回答(a)这部分，比较"检查"组和"拒绝"组里的乳房癌率不是一个好的想法。接受检查的妇女有别于拒绝接受检查的妇女。同样，比较"检查"组与对照组也不是一个好的想法，因为对照组里有两种类型的妇女——一部分人会接受，另一部分人会拒绝，你应该将整个处理组与整个对照组相比较。

(ii)研究数据表明接受检查的妇女比拒绝接受检查的妇女更富裕和受过更好的教育。穷人更多地受大多数疾病侵扰的风险、那就解释了在拒绝检查的人里出现的高死亡率。

(iii)脊髓灰质炎现场试验可以象 HIP 那样组织：(a)确定一个研究总体，比如说 1 000 000 名儿童；(b)随机分一半到处理组和一半到对照组，其中处理是指被邀请接受疫苗；(c)比较整个处理组与整个对照组的脊髓灰质炎发病率。在这个方案下，只将注射疫苗的儿童与对照组相比较是不合理的；你应该将整个处理组与整个对照组相比较。实际用于脊髓灰质炎现场试验的这个方案比较好，因为它是盲式的（第 1 章第 1 节）。

(iv)在 HIP 试验中，由于其它原因而死亡的人数是大的，并受到中等程度的机会影响；因此 837－879＝－42 这个差不是那么可靠的一个统计量。第 27 章里将有更多这方面的叙述。

(v)明式而不是盲式进行 HIP 试验似乎是个小的缺陷。

9. 不真。检查发现出业已存在的乳房癌，否则稍后也会发现，那就是检查的特征。

评论。 在处理组内，接受检查的妇女比起拒绝接受检查的妇女有诊断出乳房癌的较高偶然率。似乎有三个原因：(a)检查发现出癌病例；(b)乳房癌——象脊髓灰质炎而不象其它疾病——比起袭击穷人，它更多地袭击富人；(c)处在高风险的妇女也许更可能去接受检查。

10. 那些接触过疮疹病毒的妇女，性欲比较旺盛；所以这个依据说服力不足。（参见第 17 页例 2。）

评论。在七十年代，疮症病毒(HSV-2)被认为是引发病因。到了八十年代，来自分子生物学的新证据提出 HSV 不是最早的病因传播体，并牵涉到人类乳头状病毒的菌株(HPV-16,18)。供做参考，可参阅第 2 章的注 4。

11. 对于观察数据，这很难说。实际上，体育锻炼有时似乎引起自发性流产。若一妇女在以往的怀孕期曾经流产过——从而现时怀孕期里有更多风险——内科医生可能会告诉她减少体育锻炼。在这种情况下，体育锻炼是一种健康良好的标志，不是一种起因。

12. Rush 是位颇具偏性的观察者；他没有一个明确规定的处理组；他没有告诉我们他的病人后来发生了什么——没有跟踪；他没有对照组。

评论：正如同时期的观察者所注意到的，Rush，任凭他怎么说，他似乎对该疾病的死亡率没有影响。

13. 不真。总计，2 000 名男生中 900 人被录取，成 45%；而 1 100 名女生中 360 人被录取，成 33%。这是由于女生倾向于报考B 系，它比较难考进。参见第 4 节。

14. (a)398 中的 39 接近于 400 中的 40，成 100 中的 10，或 10%。

(b)25%　(c)25%　(d)50%。

15. (a)10%。　那散布在一个 10 000 美元的范围上，故对余下三部分，猜测每 1 000 美元范围内为约 1%。

(b)1%　(c)1%　(d)2%

第二部分　描述性统计

第 3 章　直方图

A 组　第 35 页

1. (a)2%　(b)3%　(c)4%　(d)5%　(e)15%　(f)15%
2. (a)大致相同　(b)10 000 美元和 11 000 美元之间较多
3. (a)不　(b)B　(c)20%　(d)70%
4. (a)大大地超过 50%　(b)大大地低于 50%

(c)约 50%

5. (b)班
6. 有较多人在 90—100 范围内。
7. (a)(ii). 有一个男人的峰，一个妇女的峰，而左边部分是年幼小孩的。

(b)(iii). 有男人和妇女的峰，但是没有小孩的。

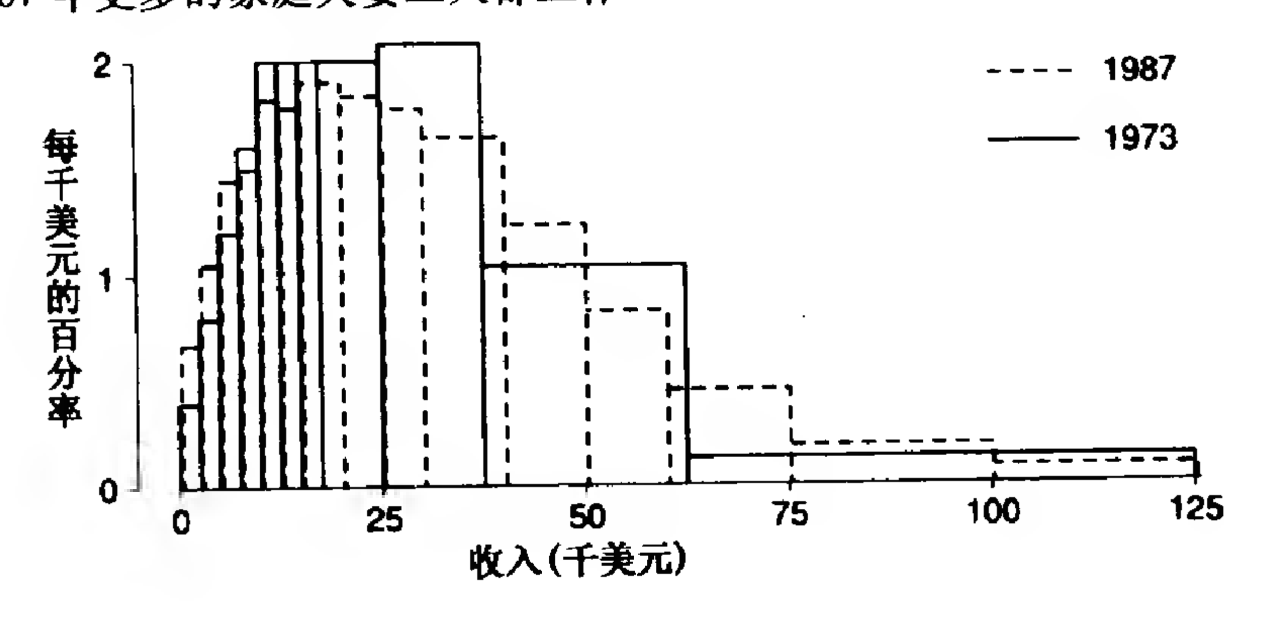
(c)(iv).

(d)(i). 汽车约为 4 英呎高，差距非常小。

直方图(v)和(vi)的单位搞错了。

8. A(ii)， B(i)， C(iii)。

9. 1973 年，一美元买的东西是 1987 年的 2.5 倍。下面的图比较了 1987 年的直方图与 1973 年的直方图——按购买力的这一变化校正。家庭收入按“公称的”美元上升了约 2.5 倍，但是按“真值的”美元(对通货膨胀校正)没有很大变化；1987 年的直方图还是证实略有更多的散布。一个主要的差别：1987 年更多的家庭夫妻二人都工作。



B 组，第 42 页

1. 1986 年的直方图在第 43 页的图 5 中给出，在那里讨论了这些峰的原因(第 44 页)。

2. 平滑 0 和 8 之间的图象。

3. 教育水平上升。比如，1986 年比 1970 年有更多的人完成中学教育并进入大学。(在本世纪中，全体居民的教育水平有了明显和稳步的上升。在 1940 年，仅全体居民的 25%完成中学教育。到了 1986 年，这百分率上升到 74%，并仍在上攀。)

4. 上升。

C 组，第 46 页

1. 每百美元 15%。

2. 答案是意见(ii)，因为(i)没有单位，(iii)的密度单位是错的。

3. (a)每支烟的 1.5%×10 支烟=15%

(b)30%　　(c)30%+20%=50%　　(d)10%　　(e)3.5%

D 组,第 49 页

1. (a)定性的。

   (b)定性的。

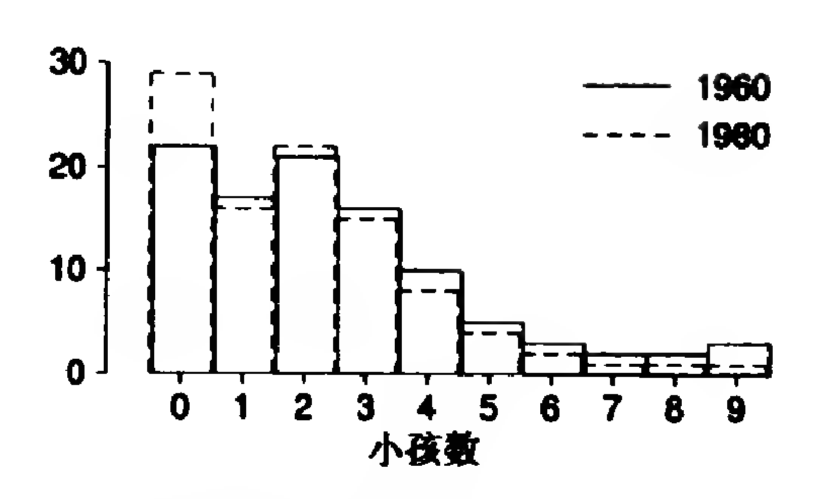
   (c)定量的,连续。

   (d)定量的,连续。

   (e)定量的,离散。

2. (a)“曾养育过的小孩数”是离散变量。

   (b)



   (c)到了 1980 年,妇女有少得多的小孩。

E 组 第 52 页

1. 总的来说,有四个小孩的母亲有较高的血压,因果关系没得到证实,存在着年龄这个混杂因子。有四个小孩母亲年龄较大。(控制年龄之后,该药物研究项目发现小孩数目与血压之间没有关联性。)

2. (i)加 10mm, (ii)加 10%。

F 组,第 54 页

1. (a)7%

   (b)5%

   (a)使用者偏于有较高的血压。

2. 口服避孕药的使用与血压增高若干毫米有关联。

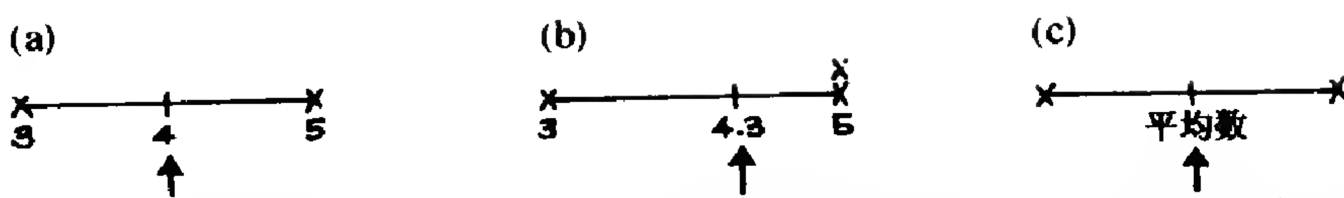
3. 较年青的妇女有略为高点的血压。

   评论。这肯定反常。大多数美国的研究结果表明收缩血压随年龄升高。通过与这些方面的其它研究相比较,避孕药研究项目里的年纪较轻的妇女有过高的血压,而年纪较大的妇女有过低的血压,这可能是多相血压测量所使用的方法里的偏性引起的。

第 4 章　平均数和标准差

A 组，第 66 页

1.



评论。对两个数，平均数在它们当中。如果你在数列中加入较大的数，平均数上升。(加入较小的数则使它下降。)平均数总是在最小和最大两数之间的某处。

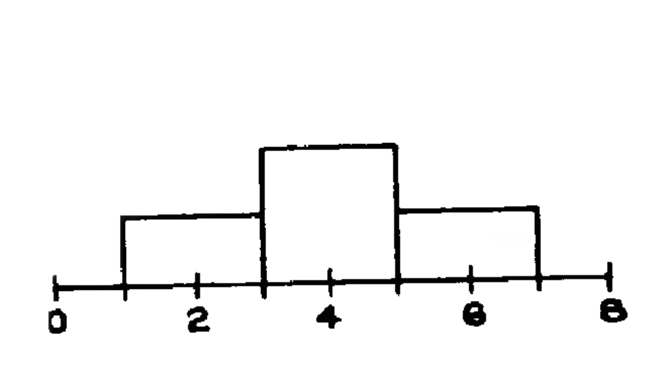
2. 若平均数是 1，这数列包含 10 个 1。若平均数是 3，这数列包含 10 个 3。平均数不可能是 4：它必定在 1 和 3 之间。

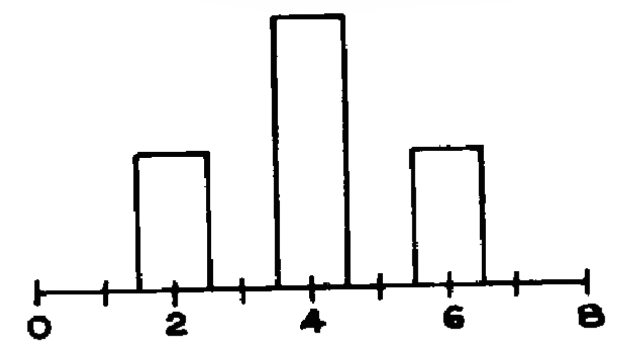
3. (ii)的平均数较大；它多了 11 这个大的项

4. (i)平均数是 2。

(ii)平均数是 3。这个数列是对数列(i)中的每一项加 1 获得；这样就对平均数加 1，并将直方图向右移 1 个单位。

(iii)平均数＝4。这个数列是将数列(i)中的每一项翻倍获得，它使平均数翻倍，并使直方图按因子 2 散布开，直方图可以按两种方式绘制如下：





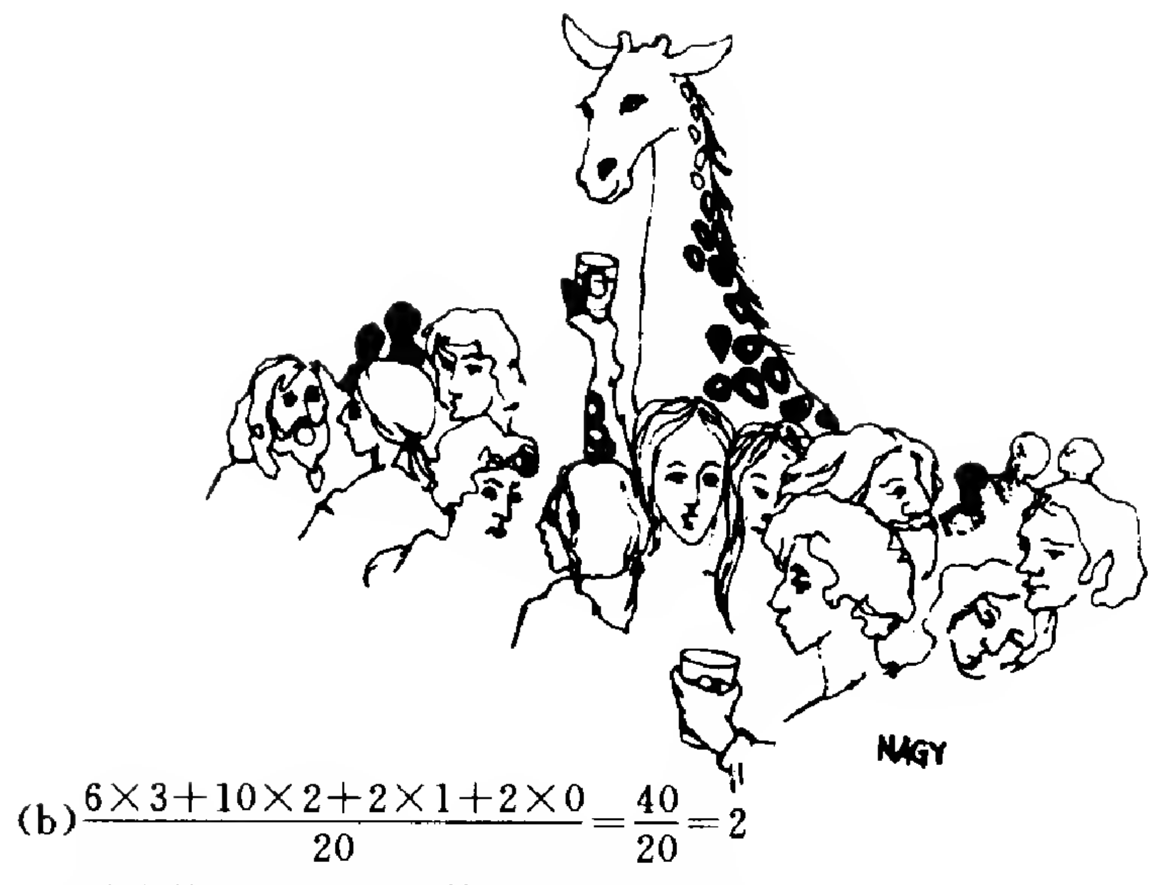
(iv)平均数是－2。这数列是改变数列(i)中每一项的符号获得，这改变了平均数的符号，并将直方图关于原点翻个个。

5. (10×66 英寸＋77 英寸)/11＝67 英寸＝5 英尺 7 英寸。或这样推理：新成员比老的平均数高出 11 英寸，故它给平均数加上 11 英寸/11＝1 英寸。

6. 5 英尺 6½英寸。当房间里的人数上升时，每个新加入的人对平均数的影响变小。

7. 5 英尺 6 英寸＋22 英寸＝7 英尺 4 英寸；它是一头长颈鹿。

8. (a) $\dfrac{3\times3+5\times2+1\times1+1\times0}{10}=\dfrac{20}{10}=2$



(b) $\frac{6\times3+10\times2+2\times1+2\times0}{20}=\frac{40}{20}=2$

(c)平均数必为 2,平均数只依赖于百分数——分布。

9. 在衰退期间,公司势必解雇职位最低,往往也是工资最低的人,这提高了(留在工资名单上的那些人的)平均工资,当衰退期结束,这些低报酬的工人重又受雇。

评论。它关系到是那些人包含在平均数里——和那些人不包含在内。

10. Rocky Mounfains(洛矶山脉)在右端,Mindanao(明达瑙)深渊在左端,而Uansas(堪萨斯)州在0(海平面)附近。

B 组,第 71 页

1. (a)50 (b)25 (c)40
2. (a)中位数=平均数 (b)中位数=平均数

   (c)平均数在中位数右边——长右尾部在起作用。
3. 20
4. 平均数必定大于中位数,故猜测为 25。(确切的答案是 27。)
5. 平均数:长右尾部
6. (a)1, (b)10, (c)5, (d)5

   (“规模”指不计符号。)

C 组,第 74 页

1. (a)平均数=0. r. m. s(均方根)大小=4

   (b)平均数=0. r. m. s(均方根)大小=10

   总的来说,数列(b)中的数在大小上较大

2. (a)10(至一位小数,确切答案是 9.0)

(b)20(至一位小数,确切答案是 19.8)

(c)1(至一位小数,确切答案是 1.3)

这些数列的平均数都是 0;均方根运算消除了符号

3. (ii)最小, (i)最大

4. 对两个数列,它都是 7;所有项有相同的大小,7。

5. r,m,s(均方根)大小为 3.2

6. r,m,s(均方根)大小为 3.1

评论。 习题 6 里的均方根大小比习题 5 中的小,有一个原因,假设我们将去比较某一个数列里的每一个数与某共同值,偏离量的均方根大小依赖于这个共同值。对某些值均方根较大,对另外一些值均方根较小。什么时候均方根最小呢? 可以从数学上证明,对平均数偏离量的均方根最小。

7. 是。总的来说误差比 3.6 大得多。如果那十名学生与其余人相象,那末计算程序有毛病。

D 组,第 77 页

1. 约 68% 在 138 到 154cm(平均数±1SD)范围内,95% 在 130 到 162cm(平均数±2SD)范围内。

2. (iii)最大;(ii)最小。

评论。所有三个数列有相同的变化范围 0 到 100。但是在数列(iii)中,更多的数远离 50,在数列(ii)中,更多的数靠近 50,比起变化范围来"散布"有更多内含。

3. (ii)最大;(i)最小

4. (a)1,由于所有数离平均数的差都是±1。

(b)2 (c)2 (d)2 (e)10

评论。SD 指出所含的项总的来说离平均数有多远;因些,只须问你自己,总的来说偏离量的大小是否接近 1,2 或 10。

5. 20 年,平均数可能是 30 年,故若答案是 5 年,则有许多人离平均数 4 个 SD;若是 50 年,则绝大多数人将在离平均数的 1 个 SD 之内。

6. 4 年,平均数约为 12 年;推理为习题 5 中,最后两个选择具错误的单位。

7. (a)(i), (b)(ii), (c)(v)

8. (a)五个 1 和五个 5 (b)五个 1 和五个 9

(c)在两个 5 之外,四个 1 和四个 9。

9. (a)全为 1;另一可能性是全为 5。

   (b)全为 1;另一可能性是全为 9。

   (c)全为 5。

10. 在试验(i)中出现了一些毛病:处理组比对照组体重重得多。(参见第 23 页上的习题 7。)

11. 它们的平均数和 SD 应该差不多相同,但是有较大样本的调查研究人员可能取到最高的男人,同样取到最矮的男人。样本越大,变化范围越大,比起变化范围,散布有更多内含。

E 组,第 80 页

1. (ii)的 SD 较大;实际上,(i)的 SD 是 1,(ii)的 SD 是 2。

2. SD 不同于平均绝对偏差,因此方法是错的。

3. 把 0 算了进去,所以是错的。

4. (a)所有三个班级有相同的平均数,50。

   (b)B 班有最大的 SD;有较多的学生远离平均数。

   (c)所有三个班级有相同的变化范围,散布比变化范围有更多内含。参见第 78 页习题 2。

5. (a)(i)的平均数=4;　偏差=−3,−1,0,1,3;　SD=2。

   (ii)的平均数=9;　偏差=−3,−1,0,1,3;　SD=2。

   (b)数列(ii)是通过对数列(i)的每一项加 5 获得的,这将 5 加到平均数上,但是,不影响对平均数的偏差,因此,它不影响 SD。将相同的一个数加到数列的每一项上不影响 SD。

6. (a)(i)的平均数=4;　偏差=−3,−1,0,1,3,　SD=2

   (ii)的平均数=12;　偏差=−9,−3,0,3,9　SD=6

   (b)数列(ii)是通过对数列(i)的每一项乘以 3 获得的。这以 3 乘平均数,它也以 3 乘离平均数的偏差,因此,SD 乘上了因子 3。以相同的一个正数乘数列的每一项时,恰好是以那个数乘 SD。

7. (a)(i)的平均数=2;　偏差=3,−6,1,−3,5;　SD=4

   (ii)的平均数=−2;　偏差=−3,6,−1,3,−5;　SD=4

   (b)数列(ii)是通过改变数列(i)中各项的符号获得的。这改变平均数的符号,以及所有离平均数的偏差的符号,但不影响 SD。

8. (a)这将使平均数增大 70 美元,但不涉及 SD。

   (b)这将使平均数和 SD 增大 5%。

9. r,m,s(均方根)大小为 17,SD 为 0。

10. SD 远小于均方根大小,参见第 67 页。

11. 不。

12. 是;譬如,数列 1,1,16 的平均数是 6,而 SD 约为 7。

13. (a)全为 2。　(b)五个 1 和五个 3。　(c)不

14. 平均数是 15,SD 是 5,它仅依赖于各种得分的百分数——分布。它不依赖于班里的学生人数(参见第 67 页,习题 8。)

15. 猜平均数,为 69 英寸。各偏离量是离平均数的偏差。离平均数的偏差的均方根恰为 SD,故你应当预期偏离 3 英寸左右。

16. 3 英吋。

第 5 章　数据的正态近似

A 组　第 91 页

1. (a)60 在平均数以上 10;那是 1 个 SD,故 60 是+1 标准单位,类似地,45 是−0.5 和 75 是+2.5。

   (b)0 对应于平均数,50。按标准单位为 1.5 的得分是平均数以上 1.5 个 SD;即平均数以上 15 分,或 65 分。得分 22 按标准单位是−2.8。

2. 平均数是 10,SD 是 2。

   (a)按标准单位,这数列是+1.5,−0.5,+0.5,−1.5,0

   (b)转换后的数列具有平均数 0 和 SD1。(常只计:当转换为标准单位时,任一数列将平均为 0 和 SD 将为 1。)

B 组,第 94 页

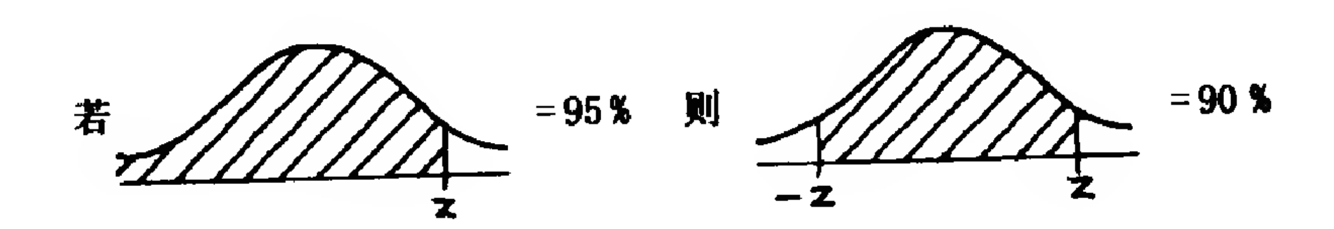
1. (a)11%　(b)34%　(c)79%　(d)80%　(e)82%

   (f)25%　(g)43%　(h)23%　(i)13%

2. (a)1　(b)1.15

3. (a)1.65

   (b)1.65;它与(a)中的 Z 相同



4. (a)100% − 39% = 61%

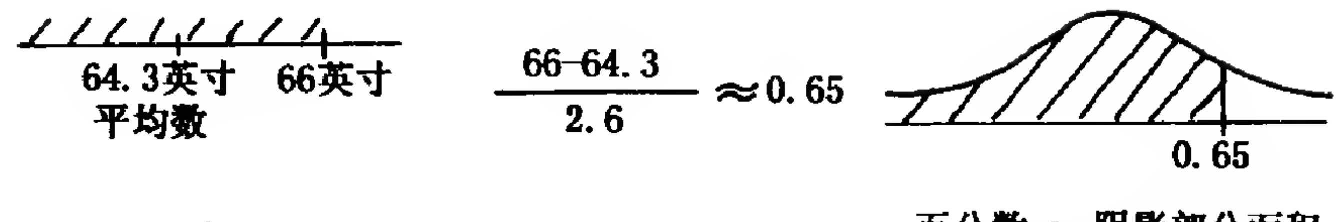
(b)在没有进一步信息下是不可能的。

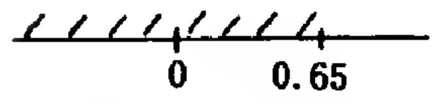
5. (a)58% ÷ 2 = 29%　　(b)50% − 29% = 21%

(c)在没有进一步信息下是不可能的。

C 组，第 98 页

1. (a)





百分数 ≈ 阴影部分面积
≈ 74%

(b)69%　　(c)1%的 0.2

2. (a)50%　　(b)62%　　(c)27%

3. (a)77%　　(b)69%

4. 精确，直方图；近似，正态曲线。

D 组，第 100 页

1. 75%。原因：90% − 10% = 80%在 8 800 美元到 68 500 美元范围内；而 10 000 美元到 70 000 美元是差不多相同的范围。

2. 7 000 美元

3. 第 25 百分位数左边的面积应为总面积的 25%，故第 25 百分位数必须相当地小于 25mm。

4 (a)它有较粗大的尾部。

(b)四分位数间距约为 15。

E 组，第 102 页

1. 她在平均数以上 2.15 个 SD，故她的百分位数序是 98%。

2. 得分是平均数以上 1.3 个 SD，它是平均数以上 1.3×100≈130 分，即 535 + 130 = 665。

3. 2.75 分——平均数以下 0.50 个 SD。

第 7 章　点和线的描绘

A 组，第 123 页

1. A = (1,2)　　B = (4,4)　　C = (5,3)　　D = (5,1)

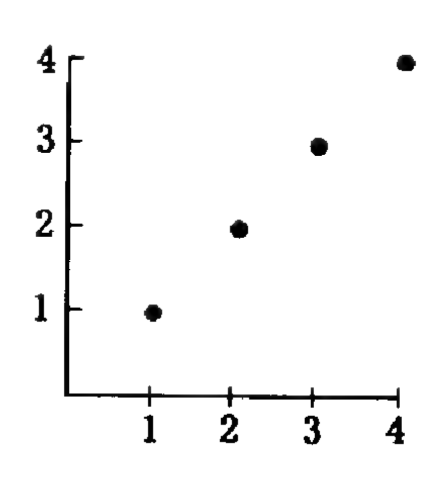
E = (3,0)

2. x 增加 3，y 增加 2。

3. 点 D。

B 组，第 124 页

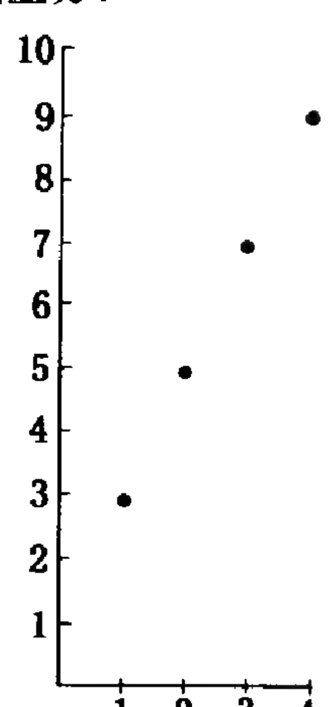
1. 四个点全在一条直线上。



2. 不在直线上的点是(1,2)，它在该直线上方。

3. 这些点全在一条直线上。

x	y
1	3
2	5
3	7
4	9



4. (1,2)在外面，(2,1)在里面。

5. (1,2)在里面，(2,1)在外面。

6. (1,2)在里面，(2,1)在外面。

C 组，第 126 页

1.

	图 16	图 17	图 18
斜率	每磅—1/4 英寸	5	1
截距	1 英寸	—10	0

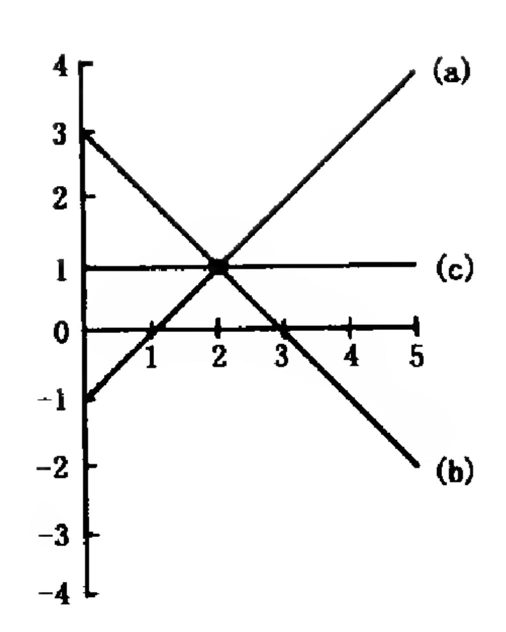
注：图 18 中，坐标轴在(2,2)处相交。

D 组，第 128 页

1. (见下页图)

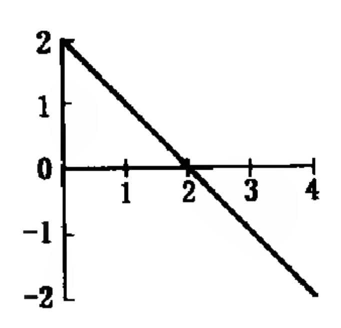
2. 在直线上。

3. 在直线上。

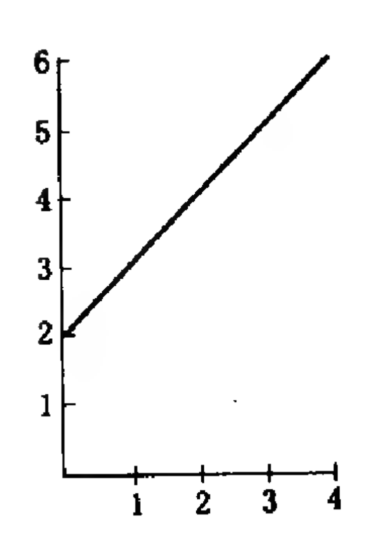


4. 在直线上方。

5.



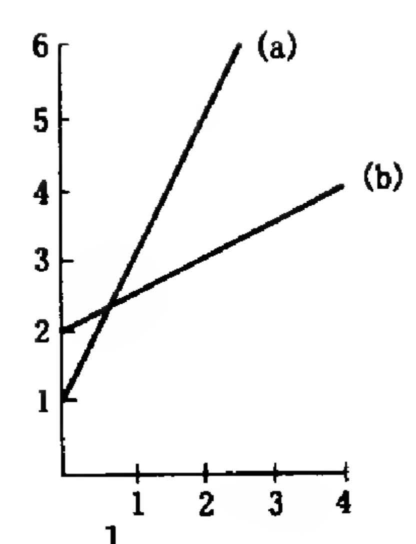
6.



E 组,第 129 页

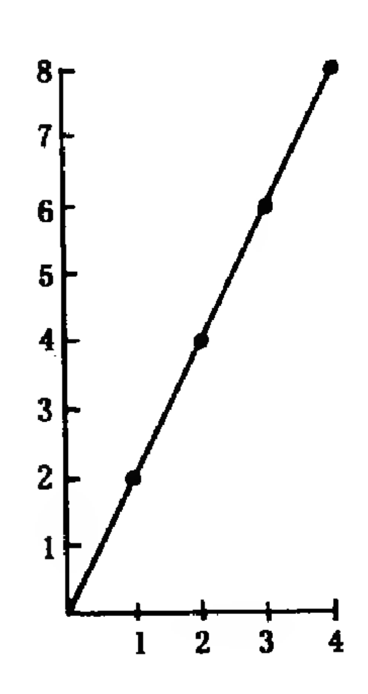
1.

	斜率	截距	在 x＝2 处的高
(a)	2	1	5
(b)	½	2	3

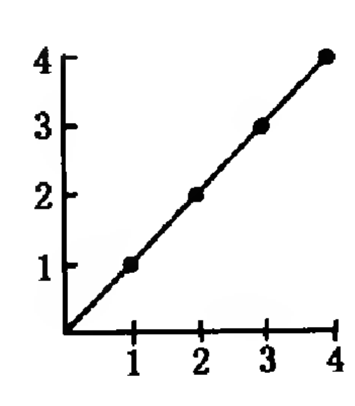


2. (a)$y=\frac{3}{4}x+1$　　(b)$y=-\frac{1}{4}x+4$　　(c)$y=-\frac{1}{2}x+2$

3. 它们全在直线 $y=2x$ 上



4. 它们全在直线 $y=x$ 上



5. (a)在直线上　(b)在直线的上方　(c)在直线的下方

6. 所有三个叙述都为真，若你理解习题 4、5 和 6，则对第Ⅱ部分，你会有条不紊。

第三部分　相关与回归

第 8 章　相关

A 组，第 137 页

1.

x	y
1	4
2	3
3	1
4	1
4	2

2. (a)x 的平均数=3　(b)y 的平均数=1.5

(c)x 的诸值

(d)

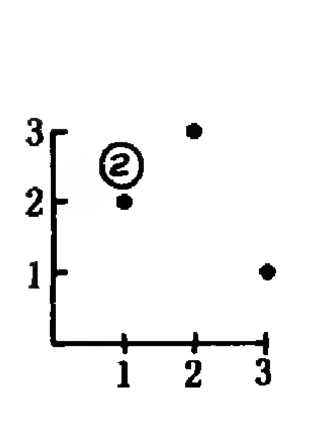
x	y
0	2
1	1
3	1
4	1
4	2
6	2

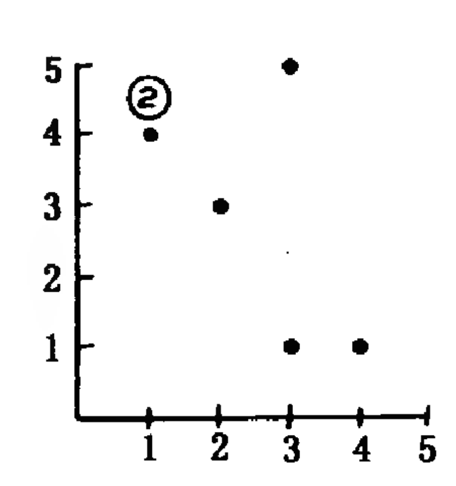
(e)x 的平均数=3,y 的平均数=1.5,x 的 SD=2,y 的 SD=0.5

3. (a)x 的平均数=1.5　(b)x 的 SD=0.5

(c)y 的平均数=2　(d)y 的 SD≈1.5

4.





5. (a)最矮的父亲,59 英寸;他的儿子,65 英寸。

(b)最高的父亲,75 英寸;他的儿子,70 英寸。

(c)76 英寸,64 英寸

(d)3 个:68 英寸,70 英寸,73 英寸。

(e)平均数=68 英寸。

(f)SD=3 英寸。

6. (a)A,B,F　　(b)C,G,H　　(c)平均数≈50

(d)SD≈25　　(e)平均数≈30　　(f)不真

(g)不真,是负相关

7. (a)75　　(b)10　　(c)20

(d)期末:期中每人得分 50 或更好。

(e)期末　　(f)真。

B 组,第 145 页

1. (a)负。车龄越老,价格越低。

(b)负。车越重,效率越低。

2. 左:x 的平均数=3.0,x 的 SD=1.0,y 的平均数=1.5,y 的 SD=0.5,正相关。

右:x 的平均数=3.0,x 的 SD=1.0,y 的平均数=1.5,y 的 SD=0.5,负相关。

3. 左图有接近于 0 的相关;它不象一条直线。

4. 相关系数约为 0.5。

5. 相关系数接近 0,心理学家称这为"稀释"。如果你限制一个变量的变化范围,通常削减相关性。

6. 散点图上的所有点都在一条斜向上的直线上,故相关系数为 1。

7. 接近 1;这就象上面一道题,有一些噪声介入数据。

评论。在 1988 年 3 月的现场人口调查中,丈夫和妻子的年龄之间的相关为 0.95。

8. (a)接近−1:你年龄越大,你越早出生;但有点模糊,关系到你的生日是在回答问卷那天之前或之后。

(b)某种正相关

9. 不真

C 组,第 147 页

1. (a)真。　(b)不真。

2. 他在高度平均数之上 1 个 SD,必重 140+20=160 磅。

3. (a)是　(b)不是　(c)是

4. 虚线

D 组,第 149 页

1. (a)x 的平均数=4,x 的 SD=2

   y 的平均数=4,y 的 SD=2

标准单位

x	y	乘积
−1.5	1.0	−1.50
−1.0	1.5	−1.50
−0.5	0.5	−0.25
0.0	0.0	0.00
0.5	−0.5	−0.25
1.0	−1.5	−1.50
1.5	−1.0	−1.50

r=乘积的平均数≈−0.93

(b)r=0.82

(c)r=−1,点全在一条斜向下直线上:y=8−x。

2. 约 50%

3. 约 25%

4. 约 5%

第 9 章　再谈相关

A 组,第 160 页

1. 最大必大于最小

2. 华盛顿与纽约较接近,并有较大相关。

3. 全年的相关较大;比如,冬天非常冷,夏天非常热。这是“稀释”的又一例(第 145 页,习题 5)

4. 不。x 与 y 间的相关与 y 与 x 间的相关相同。

5. 不。

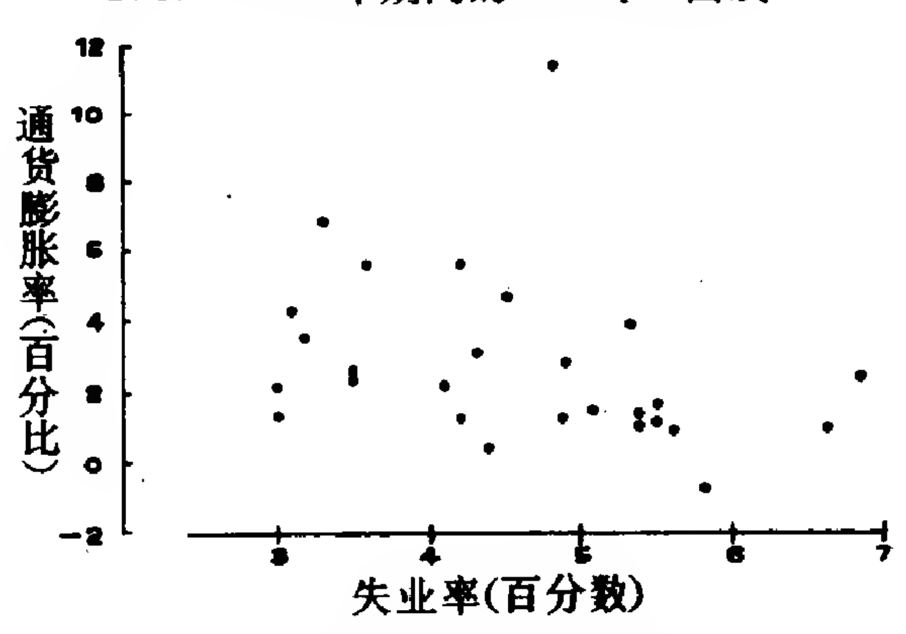
6. 不。

的相关系数。

5. 最好的解释是喝咖啡与抽烟之间的相关关系。嗜好喝咖啡的人有较大的可能是抽烟的，看来是抽烟引起心脏病

6. 这是观察研究，不是对照试验，因而在图上描上 50 年代或 70 年代的点只能弄成黑糊糊的一片。

1949—1974 年期间的 Phillips“曲线”



摘自：Economic Report of the President，1975.

第 10 章 回归

A 组，第 183 页

1. 这些男人完成了 8 年正规学校教育，这在平均数之下 5 年，他们是在平均受教育年数之下 5/3≈1.67 个 SD。估计他们的收入将在平均数之下，不过不是在它之下 1.67 个 SD——仅 r×1.67＝0.57 个 SD。用美元表示，那是 0.57×15 000 美元＝8 500 美元。他们的平均收入估计为

   总平均－8 500 美元＝21 800 美元－8 500 美元＝13 300 美元。

2. 这些男人在平均高度之上 2 英吋，或 2 英吋/3 英吋≈0.67SD。因此，他们是在平均血压之下(r 为负)，约 0.2×0.67≈0.13 个血压 SD。以 mm(毫米水银柱)表示，那是 0.13×14mm≈2mm。他们的平均血压估计为

   总平均数－2mm＝124mm－2mm＝122mm

3. (a)60　　(b)67.5　　(c)45

4. (a)70 英寸　　(b)74 英寸　　(c)68.7 英寸

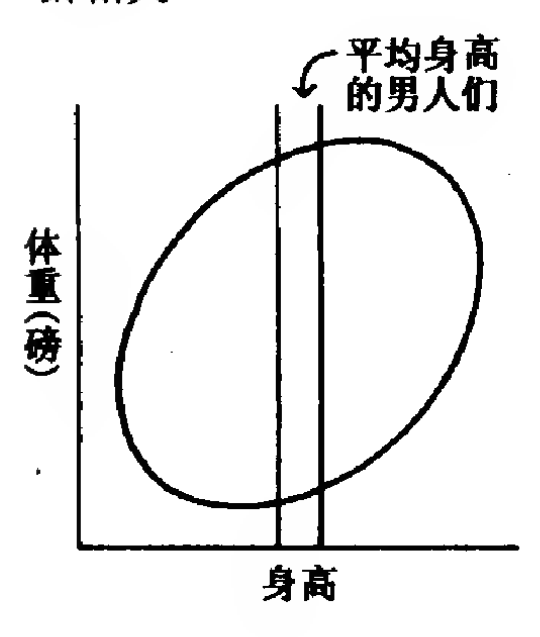
   评论。回归估计总是位于回归直线上。第 12 章将有更多叙述。

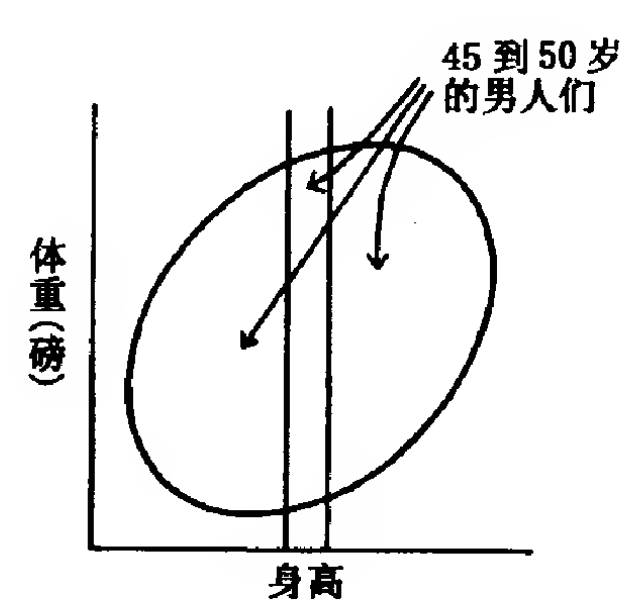
5. (a)171 磅　　(b)159 磅

   (c)－9 磅　　(d)－105 磅

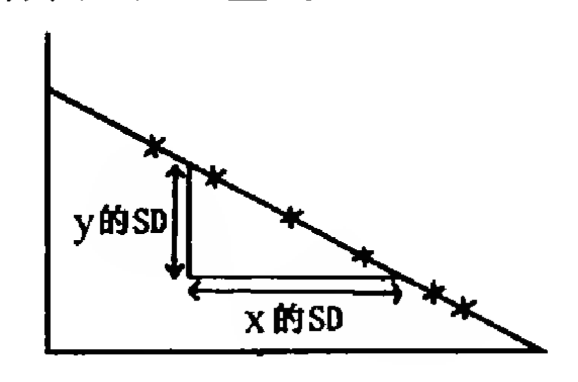
关于(c)的评论。这变得很可笑，由于公共卫生服务总署并没有收集任何 2 英吋小男人的数据，故回归直线没有对这种可能性给予太多关注。当回归直线离散点图中心越来越远时对它的信赖应该愈来愈小。

6. 不真。考虑所有男人的身高和体重的散点图。在 69 英吋上方取一纵向条形，表示身高恰好在平均数左右的所有男人，他们的平均体重应该恰好在总平均数左右。但是 45—54 岁的男人由不同的一群点表示，这些点中有一些在条形内，但大多不在，回归直线给出平均体重如何与身高而不是与年龄相关。





7. 这些点必全部位于斜向下的 SD 直线上。

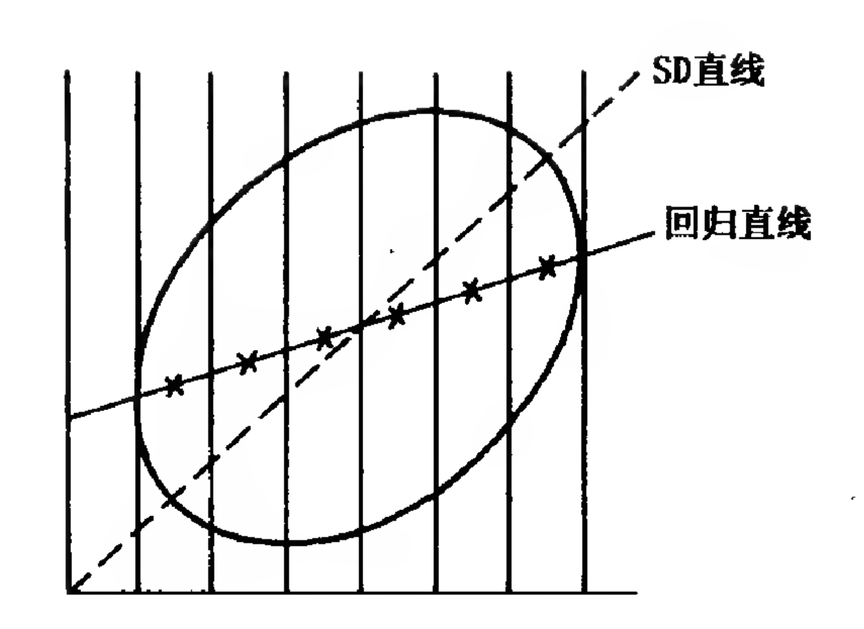


B 组，第 185 页

1. 左边两张散点图的 SD 是虚线，回归直线是实线。
右边两张散点图的 SD 线是实线，回归直线是虚线。
2. (见 690 页上图)
叉号落在实回归直线上；虚线为 SD 直线。
3. (见下页图 a、b、c、d、e、f)
评论。(f)表明回归效应在起作用——谬误场合，参见第 4 节。

C 组，第 190 页

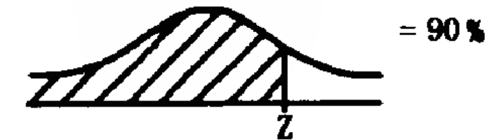
1. (a)60　(b)60　(c)67.5　(d)45

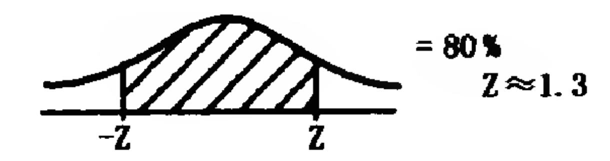


本习题是关于个体的；第 183 页上的习题 3 是关于群组的。（但算术运算相同；186—87 页）

2. (a)50%　　(b)50%　　(c)79%　　(d)38%

求解(c)：





按标准单位，他的 SAT 成绩是 1.3，他的第一年成绩的回归预测为 $0.6\times1.3\approx0.8$ 标准单位。



这对应于 79%的百分位序，在例 2 中，预测百分位序仅 69%，它接近于 50%，那是由于在例 2 中的相关系数较低。

3. 不真。确实身高在平均身高之下 0.67 个 SD 的男士，在平均意义下应该在平均体重之下 0.27 个 SD。但是，这些在这里无济于事。回归直线阐明的是平均体重如何与身高而不是与年龄相关，参见 183 页上的习题 6。

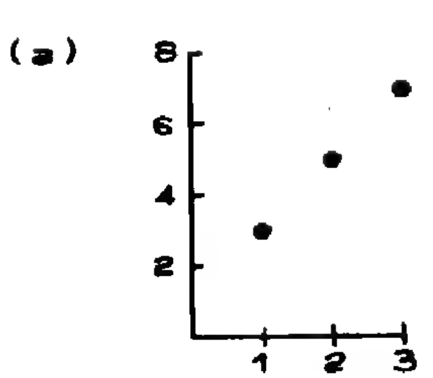
4. SD 直线——虚线。

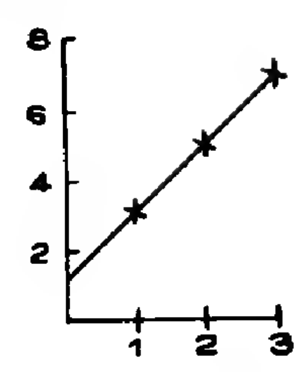
D 组，第 196 页

1. 不，它是回归效应。设想一个在两个机场进行的对照试验。在第一个机场，教官与飞行员讨论评定。在第二个机场，教官自己掌握评定结果，即便是第二个机场，两次着陆的评定也不会相同——有差别。因此出现回归效应：平均来说，低的组将有所改善，高的组将有所下降。这可能是所有空军在它们的数据中所看到的。

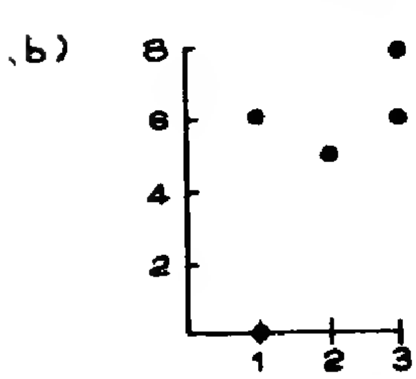
散点图	平均数与回归直线图形

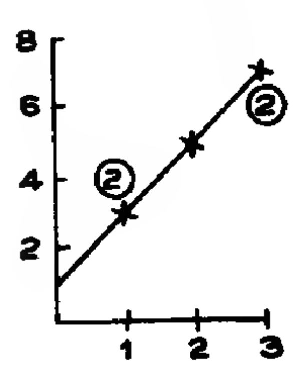
(a)



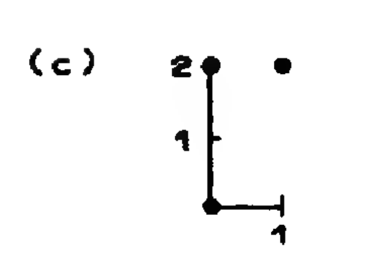


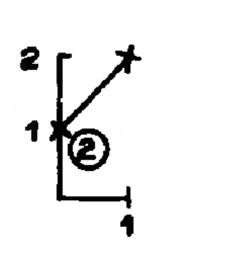
(b)



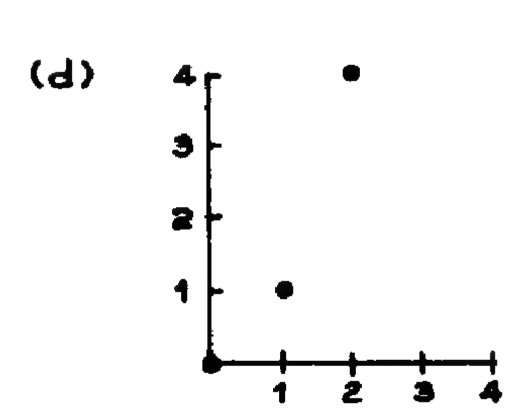


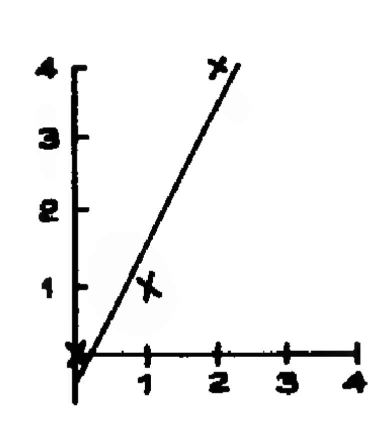
(c)



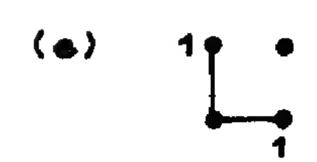


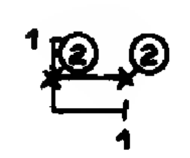
(d)



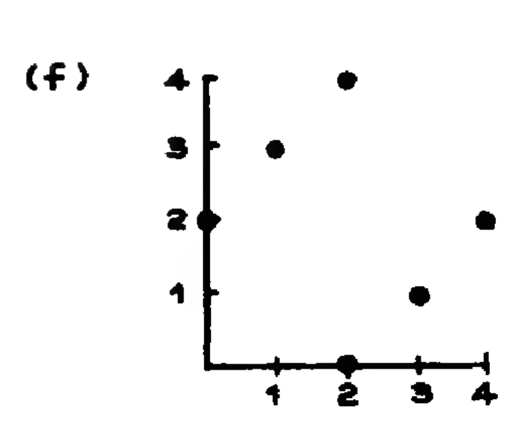


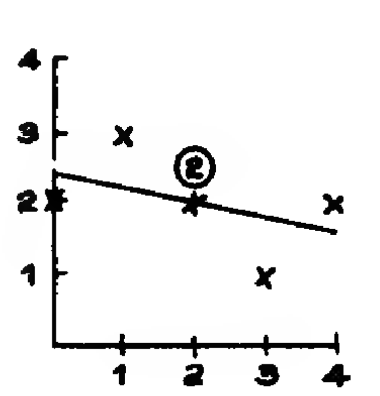
(e)





(f)





2. 不，它看似特别指导的效应——回归只是使他们接近平均数，但他们到了另一边。

3. 50%以下许多——回归效应。

评论。云状图更形似一只冰淇淋蛋筒，不象橄榄球(215 页上图 10)，但相同的原则仍适合。

4. 61 英寸父亲的儿子平均比 62 英寸父亲的儿子高些，这只是机会变差。由于抽取的侥幸，Peavsan 取到了过多的家庭，其父身高 61 英寸而儿子特高。

5. (a)68 英寸。　(b)69 英寸，2 英寸。

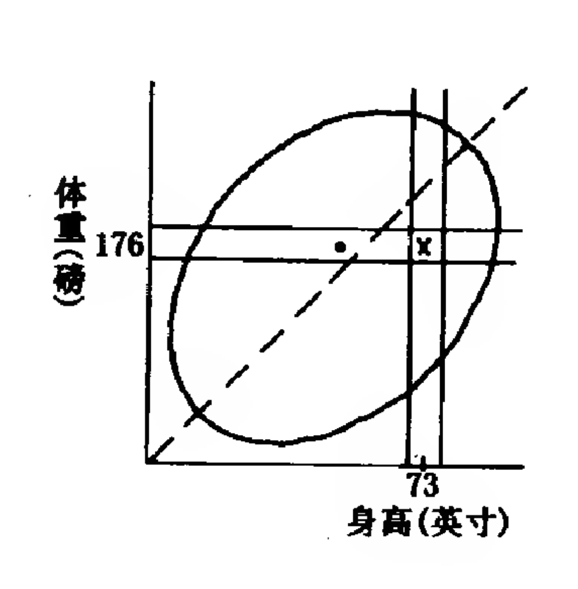
(c)单个的人不趋于平均：68 英寸高的父亲，他们的儿子的高度有很大的散度。但是，如果你在某次检测中取一个极端的组(高或矮)，在第二次检测时他们的平均数将接近总平均数。

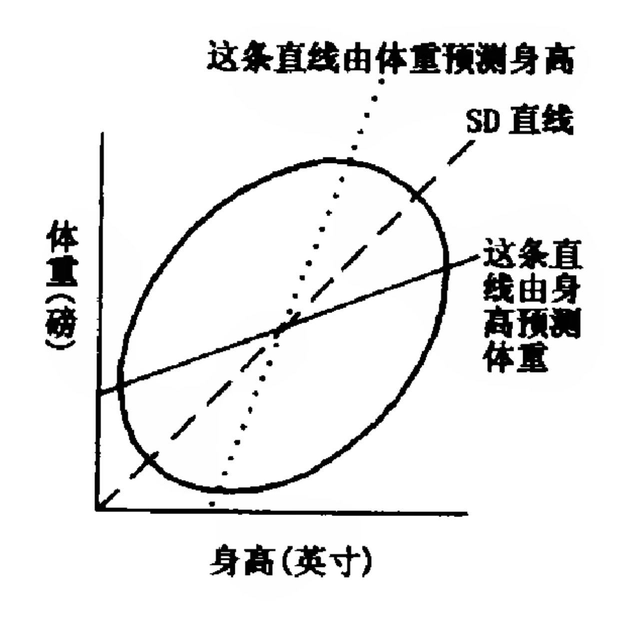
E 组，第 199 页

1. 不真，这里有两个完全不同的男人组。(参见下面的散点图)73 英寸高的人在纵向条形中，他们的平均体重约 176 磅，如叉号所示。体重 176 磅的人在横向条形中，他们平均身高由黑点表示，它比 73 英寸小得多。

记住，存在两条回归直线

- 一条是体重关于身高的。
- 一条是身高关于体重的。





2. 不真，父亲平均仅 69 英寸；必须用另一条回归直线。

3. 不真，这恰如习题 1 和 2，(位处第一年考试 69%分位数的学生的典型代表应位处 SAT58%分位数；采用另一条回归直线。)

第 11 章　回归的均方根误差

A 组，第 207 页

1. (a)0.2　(b)1　(c)5

2. 几千美元

3. 应采用均方根误差较小的一条，因为总体来说将更为精确。

4. (a)8 分——一个均方根误差

(b)16 分——二个均方根误差。

B 组，第 209 页

1. $\sqrt{1-(0.6)^2}\times 10=8$ 分。

2. (a)用平均数猜，65。

(b)10。若你用回归直线，均方根误差由公式(习题 1)给出，若你用平均数，均方根误差为 SD(参见 80 页和 82 页上的习题 15—16)

(c)采用回归直线，且由公式均方根误差为 8 分(习题 1)。

C 组，第 212 页

1. (a)(iii)　(b)(ii)　(c)(i)

2. (a,b)　x 的平均数≈4，SD≈1，y 的平均数≈4，SD≈1

(c)r≈0.8　(d)0　(e)0.6　(f)4.8　(g)0.6

D 组，第 216 页

1. (a)真

(b)真；散点图是等方差性的，因此，在各纵向条形中，实验对象偏离回归直线相同的量。

(c)不真，因为散点图是异方差性的；9 分是平均偏离量的一种，但预测误差随高分而变得更大。

2. (a)$\sqrt{1-(0.5)^2}\times 2.7\approx 2.3$ 英寸

(b)71 英寸——回归方法

(c)2.3 英寸。散点图是等方差性的，因此，对任何父亲的身高，儿子的身高偏离回归直线相同的量，这个偏离的量为该回归直线的均方根误差。

(d)预测值为 68 英寸，且它可能偏离 2.3 英寸左右。

3. (a)$\sqrt{1-(0.4)^2}\times 10\ 500$ 美元≈9 600 美元。

(b)14 000 美元——回归方法。

(c)这不能由所给的信息决定。9 600 美元是平均偏离回归直线的量的一

种。但散点图是异方差性的，故不同条形中偏离回归直线的量也不同，受较高教育的人其收入的散度也较大，故偏离量大于 9 600 美元。

(d)预测值为 9 600 美元。偏离量不能确定，但它将小于 9 600 美元。

4. (a)68 英寸，平均数

(b)3 英寸，SD。

(c)回归，若双胞胎中的一个是 6 英尺 6 英寸，猜另一个为 6 英尺 5½英寸。

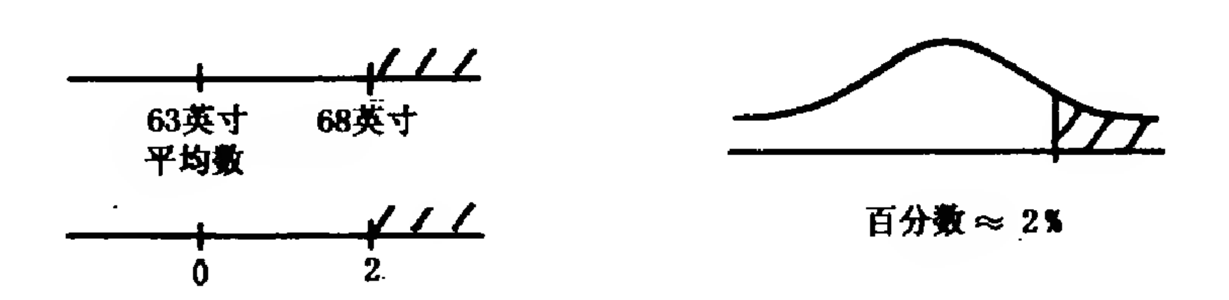
(d)$\sqrt{1-(0.95)^2}\times 3\approx 0.9$ 英寸

评论。(i)若 r=1，你应该猜双胞胎中第二人的身高与第一人相同。但 r 略小于 1，因此你将第二人的身高向平均数方向少许退回一些。

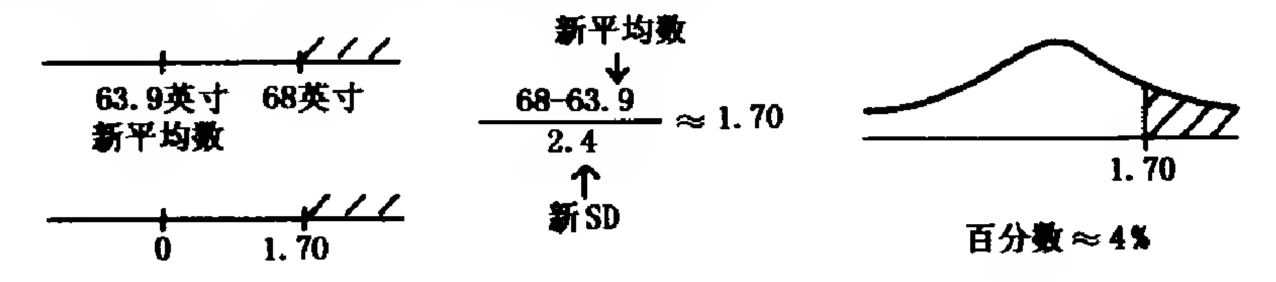
(ii)对(d)的答案大大小于对(b)的答案。当 r=0.95 时，当你采用回归直线时，均方根误差存在较大的缩减。

E 组，第 219 页

1. (a)



(b)新的平均数≈63.9 英寸，新的 SD≈2.4 英寸



2. (a)14%　　(b)33%

3. (a)38%　　(b)60%

第 12 章　回归直线

A 组，第 231 页

1. (a)1 400 美元 ×8+2 200=13 400 美元

(b)1 400 美元×12+2 200 美元=19 000 美元

(c)1 400 美元×16+2 200 美元=24 600 美元

2. (a)240 盎司=15 磅

(b)20 盎司

(c)3 盎司氮肥生产 18 磅 12 盎司稻谷，4 盎司氮肥生产 20 磅稻谷。

(d)对照试验。

(e)是，该回归直线拟合得十分好(r=0.95)，并且 3 盎司接近于使用的值。

(f)不。那离所使用的量太远。

3. 白人：29 600 美元。黑人：23 500 美元

评论. 白人和黑人的收入分布不对受教育程度进行调整的话，则难于互相比较，回归亦如此。

4. 对受教育 12 年的人，身高预测为 69.75 英寸；对 16 年的人，身高预测为 70.75 英寸，上大学显然并不影响身高。这项观察研究获得了由家庭背景中的某第三个因素引起的身高与受教育程度之间的相关。

5. 预测儿子的身高=0.5×父亲的身高+35 英寸

6. 预测父亲的身高=0.5×儿子的身高+33.5 英寸

评论. 存在两条回归直线：一条由父亲的身高预测儿子的身高，另一条由儿子的身高预测父亲的身高(第 10 章第 5 节)。

B 组，第 238 页

1. 根据第 234 页，回归方程为

预测长度=m×重物重量+b

=(0.05 厘米/公斤)×重物重量+439.01 厘米

将重物重量以 3 公斤和 7 公斤代入得长度的估计为 439.16 厘米和 439.36 厘米，这方法不可用于重物重量为 50 公斤场合，因 50 公斤比表 1 里的任一重量大得多，金属丝可能突然折断。

2. 相关系数为 0.996 854，方程为

预测消费=0.9 135×收入−67 美元。

这是 Keynes 的消费函数的简单形式。

第四部分 概率

第 13 章 什么是机会？

A 组，第 248 页

1. (a)(vi)　(b)(iii)　(c)(iv)

(d)(i)　(e)(ii)　(f)(v)　(g)(vi)

2. 约 500。

3. 约 1 000

4. 约 14

B 组，第 251 页

1. 约 300 只红的，200 只蓝的。

2. (a)308 美元(b)273 美元(c)300 美元

3. 300 美元－200 美元＝100 美元

4. 300×8 美元－200×10 美元＝400 美元

5. 盒子(ii)，因为[3]比[2]支付得更多，另一张票是相同的。

6. (a)25%　(b)10%

C 组，第 252 页

1. (a)1/4；见例 2 的(a)。
   (b)1/3；抽出[2]后剩下 3 张票子。

2. (a)1/4　(b)1/4
   以有放回方式，盒子保持不变。

3. (a)1/6
   (b)1/6——这跟第一次发生的是什么没有关系。

4. (a)1/2　(b)1/2

5. (a)1/52　(b)1/52　(c)13/52＝1/4
   (d)9/48：剩下 48 张牌，其中 9 张是梅花

D 组，第 255 页

1. 1/6×1/6＝1/36≈3%

2. 4/52×4/51×4/50≈5/10 000

3. $(1/2)^3=1/8$

4. “至少一个么点”是较佳选择：你得选择必须做对六道题中的至少一道这种考试，而不选必须全做对六道题的考试。

5. 硬币必须出现“背面，头像”；机会是 1/4。

6. 不真，计算假设妇女的百分数在所有年龄组都相同，而并不如此：妇女比男士活得更长。

E 组，第 258 页

1. (a)独立　(b)独立　(c)相依

2. (a)1/8

(b)1－1/8＝7/8

(c)7/8；如果你没有得到 3 次头像，你至少得到一次背面；故(b)和(c)是一样的。

(d)7/8；只要把(c)中的头像和背面位置调一调则可。

3. 十年是 520 个星期，故机会是

$(999\ 999/1\ 000\ 000)^{520} \approx 0.9\ 995$

评论. 在 New York State Lotto(纽约州抽奖)中，你赢得某东西的机会约为 1/12 000 000。

4. 这不真。它就象你找不到体温表就说某人没有体温一样。要判断两个事物是否独立时，你假装知道第一个的结果，然后查看第二个的机会是否变化。强调的是"假装"一词。

5. (a)5%　　(b)20%

为了算出(a)的结果，假设在班里你有 80 个男士和 20 个女士。你亦有 15 张标着"新生"的牌和 85 张标着"二年级生"的牌。你要给每名学生发一张牌，使得尽可能少的女士得到"二年级生"。最佳策略是给每位男士发一张二年级生的牌；剩下 5 张必须发给女士，另 15 张新生牌发给其余 15 名女士。

评论. 若年级和性别相互独立，二年级女士的百分数将为 20%的 85%＝17%，在两个极端之间。

6. 若实验对象从小的一堆里抽得黑桃 A，他有 52 分之 13 的机会从大的一副纸牌里抽得一张黑桃，并赢得奖品。如果他抽得梅花二或任何其它的牌也一样。故答案是 13/52＝1/4。

第 14 章　再谈机会

A 组，第 268 页

1. ⚀⚄　⚃⚁　⚁⚃　⚄⚀.　机会是 4/36。

2. 最常出现，7：最少出现，2，12。(象习题 1 中那样，利用图 1 求得各总和的机会。)

3. 存在 25 个可能结果；机会是 5/25。

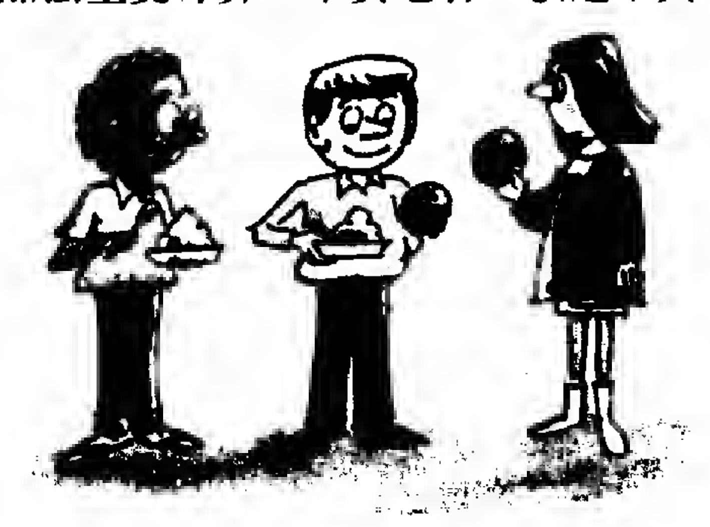
B 组，第 271 页

1. 首先要做的事是理解问题的含意。对于统计学家，这项叙述是有关吃饼干或吃冰淇淋的小孩人数，其中包括两样都吃的贪吃者。这个数目依赖于小

孩们所做的选择，下表列出两种可能情况。

	只吃饼干	只吃冰淇淋	两样都吃	两者都不吃
A	12	17	0	21
B	3	8	9	30

在第一种情况下，12 名儿童只吃饼干，17 名只吃冰淇淋，没人两样都吃，和 21 名两样都不吃。故 12+17=29 名吃饼干或冰淇林。第二行表示另一种可能，其中 9 名儿童饼干和冰淇淋两样都吃。在这种情况下，吃饼干或冰淇淋的人数是 3+8+9=20。仅供做检查：吃饼干的人数为 3+9=12，而吃冰淇淋的人数是 8+9=17，就象问题中所给。但是吃饼干或冰淇 淋的人数不是 12+17，因为这种加法重复计算 9 个贪吃者。叙述不真。



2. 它们相同。
3. 不真。简单地把两个机会相加重复计算了⚀⚁的机会。
4. 不真。在任一特定抽取中得到[7]的机会是 10 分之 1，但是这些事件不是互不相容的。
5. 真，100%－(10%+20%)=70%。
6. (a)不真；1/2×1/3=1/6，但 A 和 B 可能相依：你需要给定 A 下 B 的条件概率。

   (b)真；参见第 13 章的第 5 节。

   (c)不，机会为 0。

   (d)不真；1/2+1/3=5/6，但你不可以把两个概率相加，因为 A 和 B 可能不是互不相容的。

   (e)不真；如果它们相互独立，它们就有同时发生的某些机会，所以它们不可能是互不相容的。由于它们不是互不相容的，你不可以把它们的机会相加。

(f)真

7. 2.1%+4.6%=6.7%。如果你有三张牌是同花，你不可能有两对：不相容事件。

8. (a,b)真；参见第 13 章中的例 2。

(c)不真，“最上面一张牌是梅花 J”和“最底下一张牌是方块 J”不是互不相容的，故你不能把机会相加。

(d)真，“最上面一张牌是梅花 J”和“最底下一张牌是梅花 J”是互不相容的。

9. (a)1/52 的参赛者向前跨。

(b)1/52 的参赛者向前跨；第 13 章中的例 2。

(c)第一次得红心 A，第二次得红心 K 者向前跨两次。（在赢得周末方面，那是两次命中。）向前跨两次的比率是 1/52×1/52。

(d)不真；如(c)所示，这两个事件不是互不相容的，故相加重复计算了同时发生的机会。机会为

$$1/52+1/52-1/52\times 1/52$$

10. (a)1/52 的参赛者向前跨。

(b)1/52 的参赛者向前跨。

(c)如果你第一次得到红心 A，第二次就不可能得到它；无人向前跨两次。

(d)真；如(c)所示，这两个事件是互不相容的，故相加是合理的。

C 组，第 276 页

1. (a)3/4 (b)3/4 (c)9/16 (d)9/16(e)7/16

2. (a)无 A 的机会 $=(5/6)^3\approx 58\%$，故至少一个 A 的机会≈42%，象 de Méré 那样，以掷 3 次代替 4 次。

(b)67% (c)89%

3. $1-(35/36)^{36}\approx 64\%$。

4. 抛 22 次中 17 点不出现的机会是 $(31/32)^{22}\approx 49.7\%$。因此，在抛 22 次中它出现的机会为 100%−49.7%=50.3%。故这种赌博（赌偶数）亦有利于球的庄家。不幸的投机。

5. 50 次战斗任务幸免于难的机会是 $(0.98)^{50}\approx 36\%$。

第 15 章　二项系数

A 组，第 285 页

1. 这数是 4。

2. 这数是 6.

3. (a)$(5/6)^4=625/1\ 296\approx 48\%$

(b)$4(1/6)(5/6)=500/1\ 296\approx 39\%$

(c)$6(1/6)^2(5/6)^2=150/1\ 296\approx 12\%$

(d)$4(1/6)^3(5/6)=20/1\ 296\approx 1.5\%$

(e)$(1/6)^4=1/1\ 296\approx 1\%$的 0.8

(f)加法法则:$(150+20+1)/1\ 296\approx 13\%$

4. 这与习题 3(a－c)相同:考虑 1 为红,2 至 6 为绿,更明确些,设想两个人,A 和 B,进行不同的机会实验:

   · A 掷一只骰子 4 次并计数 A 的次数。

   · B 从盒子

   R G G G G G

   随机有放回地抽取 4 次并计数 R 的次数。

   工具不同,但是只要关心的是得到红的任一特定次数的机会,则这两个实验是等价的。

   · 掷四次就象抽四次一样。

   · 掷的结果是独立的,抽的也如此。

   · 每次掷有 6 分之 1 的机会使(得 A)计数增加 1;对每次抽(得红)类同。

5. 正好得到 5 次头像的机会是$\frac{10!}{5!\ 5!}(\frac{1}{2})^{10}=\frac{252}{1\ 024}\approx 25\%$。正好得到 4 次头像的机会是$\frac{10!}{4!\ 6!}(\frac{1}{2})^{10}=\frac{210}{1\ 024}\approx 21\%$。正好得到 6 次头像的机会与此相同。由加法法则,得到 4 到 6 次头像的机会是 $672/1\ 024\approx 66\%$。

6. $4(1/2)^4=25\%$

7. $\frac{10!}{3!\ 7!}(\frac{1}{6})^3(\frac{5}{6})^7\approx 16\%$。

第五部分　机会变异

第 16 章　平均数法律

A 组,第 298 页

1. 误差按绝对值为 50,按自分数为 5%。

2. 误差按绝对值为 1 000,按百分数为 1%的 1/10,将这与上题相比较,机会

误差按绝对值上升(中 50 上升到 1 000)但按百分数下降(由 5%降升 1% 的 1/10)。

3. 不真,机会保持在 50%。见第 248 页。

4. 抛 10 次,当抛的次数增加时,头像的次数正好等于期望次数的机会越来越小——存在更多的可能结果。

5. 抛一百次:当抛的次数上升,头像的百分数可能更接近 50%。(抛 10 次,得到 4 到 6 次头像的机会约为 66%:参见 241 页上的习题 5。抛 100 次,得到 40 到 60 次头像的机会约为 95%,就象第 18 章中将看到的那样。)

6. 抛 10 次,它是习题 5 的对立面。

7. 选择(i)较好,恰如上题。

8. 选择(ii),原因是机会误差。

9. 有或不放回几乎相同。

10. 相同,两者都有 50%的[−1]和 50%的[+1]

11. 最后机会误差将变大和取负值。而后,它又会取正值,摆幅越来越宽。

B 组,第 302 页

1. $47\times1+53\times2=153$

2. (a)100,200 (b)50,50 (c)$50\times1+50\times2=150$

3. (a)100,900

(b)$33\times1+33\times2+33\times9\approx400$

评论. 400 不是 100 和 900 之间的半当中

4. 在所有三种情况下都猜 500;(iii)最好,(i)最差。

5. 选择(ii)削减变异,因为盒子变得较小,故(ii)较佳。在两种情况下都猜 50。

6. 盒子(i)较佳:它有较少的−1 和相同的叉。

7. 选择(i)和(ii)可行。你的净收益是你赢的数目和输的数目的总和,注意符号:赢是正的,输是负的。

C 组,第 307 页

1. (i)与(ii)相同,(iii)意指所有十个抽得数必须为"1",它比(i)差。

2. "0"的机会是 10 分之 1;"小于等于 3"的机会是 10 分之 4;"大于等于 4"的机会是 10 分之 6。若心有疑虑,参阅第 13－14 章。

3. 选择(i)不适合;抽得数之和与净收益无关。选择(ii)不适合,它称你每赌一次赢 17 美元的机会是 36 分之 2,但你的机会是 38 分之 2。选择(iii)正确,若心有疑虑,复习一下 257 页上的例 1

4. 你的净收益等同于随机有放回地从盒子

1 张 \$36 的票子 215 张 −\$1 的票

取出的 10 个抽得数的和。这是很糟的游戏。

5. 取自盒子 | 0 | 1 | 的十个抽得数。

第 17 章 期望值与标准误差

A 组，第 314 页

1. (a)$100\times2=200$ (b)-25 (c)0 (d)662/3

评论(d)."期望值"不必是可能取值中的一个。它就象说平均家庭有 2.1 个小孩。这是有意义的，虽然"平均家庭"是统计的虚构。

2. 这与取自盒子 | 1 | 2 | 3 | 4 | 5 | 6 | 的二个抽得数和的期望值相同。故答案是 $2\times3.5=7$ 个正方形。

3. 模型在 258 页上给出。盒里数的平均数是

(35 美元−37 美元)/38=−2 美元/38≈−0.05 美元

期望净收益为 100×(−0.05 美元)=−5 美元。你可预期输掉 5 美元左右。

4. 盒子在 304 页上。盒子的平均数是

(18 美元−20 美元)/18=−2 美元/38=−0.05 美元

故期望净收益为 100×(−0.05 美元)=−5 美元

评论. 习题 3 和 4 指出对任一种赌法(赌数或赌红和黑)，你可预期每赌一次你会输掉赌注的 1/19。

5. −50 美元. 寓意：你赌的次数越多，就输得越多。

6. 盒子的平均数是(18x−20 美元)/38。要使它公平，这必须等于 0。方程为 18x−20 美元=0。故 x≈1.11 美元，他们应该付给你 1.11 美元。

7. 就象冒险家所想的，球的主人应支付 31 磅。寓意：冒险家可能找到了乐趣，但是得益的是球的主人。

B 组，第 318 页

1. (a)盒子的平均数是 4；SD 是 2。故和的期望值为 $100\times4=400$；和的 SE 为 $\sqrt{100}\times2=20$

(b)猜 400，偏离 20 左右。

2. 净收益等同于取自盒子 | −\$1 | \$1 | 的 100 个抽得数之和。盒子的平均数

为 0;SD 为 1 美元,100 个抽得数的期望值为 0 美元;和的 SE 为$\sqrt{100}\times 1$美元=10 美元。故你的净收益约为 0 美元,加减 10 美元左右。

3. 对选择(ii),所有的数都太接近 50;无一数偏离 5 以上,对选择(iii),所有的数过份有规律地交替。选择(i)为所选。

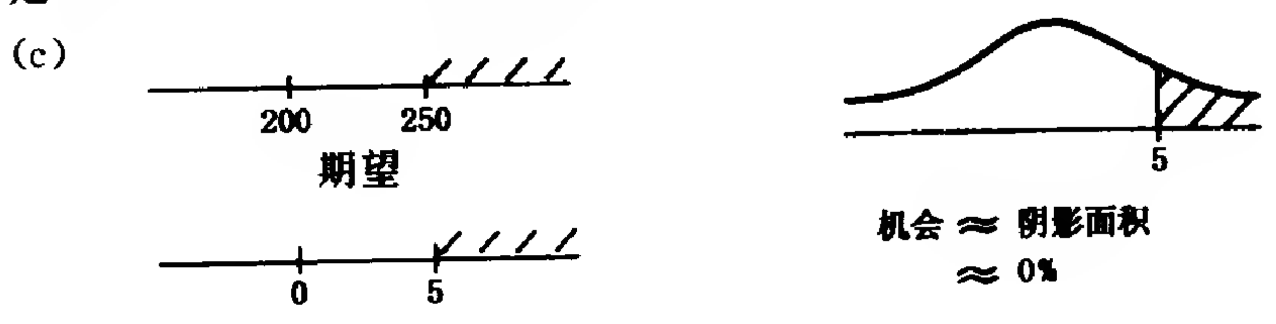
4. 是。机会是小的,但是正数,只要你等待足够长,小概率事件也会发生。

C 组,第 322 页

1. (a)最小,100;最大,400

(b)盒子的平均数是 2,SD 是 1。和有一个 100×2=200 的期望值;和的 SE 是$\sqrt{100}\times 1=10$。和将约为 200,加减 10 左右。

(c)



2. (a)最大,900;最小,100。

(b)机会≈68%

3. (a)期望值为 0,故和约为 0,你的最佳希望是机会变差,这随抽取次数而增大;选 100。

(b)与(a)同。

(c)这时机会变差于你不利—— 选 10。

4. (i)和的期望值=500,和的 SE=30

(ii)和的期望值=500,和的 SE≈28

两个和都在 500 左右,但和(i)将离得更远些,在(a)和(b)中,机会变差有益____选(i)。在(c)中机会变差有害____选(ii)

5. 95%

6. 无论他们赢 25 000 美元(具有机会 20/38≈53%)或输 25 000 美元(具有机会 18/38≈47%)。答案是 50%。

评论.相比之下,赌场更欢迎大量的小赌注,不那么欢迎一笔大赌注,因为前者得益几乎是保证的,后者存在大量风险。

7. 一个数将有 35 美元的收益,其余 37 个数将输,因此该赌徒必输 2 美元。

评论.赌场喜欢赌徒们分散他们所下的赌注。

8. 不真。该公式仅适用于从一只盒子随机有放回地抽取。这人将盒子平分为

二，这样做将减小变异。

D 组，第 325 页

1. (a)不，用 7－(－2)＝9 代替 5。　(b)是　(c)是。

   (d)不____这列数给出 3 个不同的数，故该捷径不适用。

2. 0.9 986 美元，这容易用捷径方法来做。

3. 净收益等同于取自盒子

   | $2 | −$1 | −$1 | −$1 |

   中 100 个抽得数之和，盒子的平均数为

   (2 美元－1 美元－1 美元－1 美元)/4＝－0.25 美元

   SD 为[2 美元－(－1 美元)]×$\sqrt{1/4\times3/4}\approx1.30$ 美元

   赌 100 次中的净收益将约为 100×(－0.25 美元)＝－25 美元，加减 $\sqrt{100}\times1.30$ 美元＝13 美元左右。

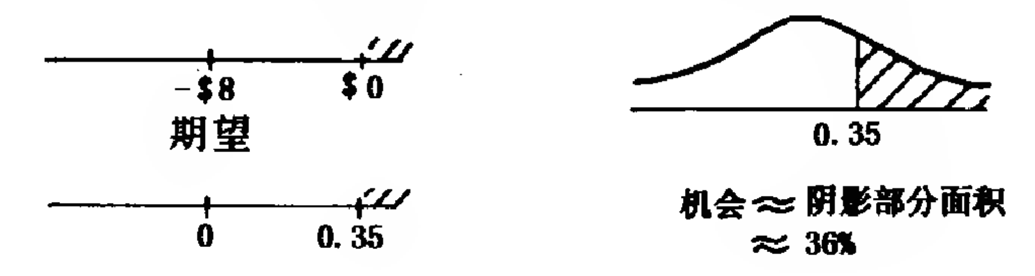
4. (a)根据庄家的观点，在庄家专号上下一美元赌注等同于从盒子 | 5 张 −$6　33 张 +$1 | 中抽取一次，盒子的平均数为[5×(－6 美元)＋33×1 美元]/38≈0.08 美元，故每一美元赌注庄家预期可赚 8 美分，就庄家来说，这是有利的赌法。

   (b)下注者的净收益等同于随机有放回地从盒子

   | 5 张 +$6，33 张 −$1 |

   中取出的 100 个抽得数之和。盒子的平均数为－0.08 美元；SD 为[6 美元－(－1 美元)]×$\sqrt{5/38\times33/38}\approx2.37$ 美元。

   下注者 100 次下赌中的净收益为－8 美元，加减 24 美元左右。



5. 在某 Seclivn 上下注一美元 100 次的期望净收益是－5 美元；SE 为 14 美元。在红色上下注 100 次的期望净收益是－5 美元；SE 为 10 美元。选择(i)和(ii)有相同的期望净收益。但(i)的 SE 较大，即更具变异性，故(a)不

真,(b)和(c)为真。

E 组,第 331 页

1. (a)你不可能把字相加,故盒子(i)被排除,对盒子 2(iii),你每次有 3 分之 2 的机会增长,而应该只有 2 分之 1 的机会。盒子(ii)是所需的。

   (b)盒子的平均数=0.5,同样盒子的 SD=0.5。16 个抽得数之和的期望值为 16×0.5=8;SE 为$\sqrt{16}\times 0.5=2$。头像的次数将约为 8,加减 2 左右。

2. (a)不真——公式要求用分数,不是数,SD 为$\sqrt{2/5\times 3/5}$

   (b)不真——乘以 2。　　　　(c)真。

3. 新盒子:| 0 | 0 | 0 | 0 | 1 |。它是±3SE,机会约为 99.7%。

4. 新盒子:| 0 | 1 |。它是 1SE 或更多,机会约为 16%。

5. 预期约 68,考虑 69。

6. 约 99.7%;参见 329 页上的例 5。

7. 约 99.7%。

   评论. 比较习题 6 和 7,当抛硬币的次数从 10 000 增加到 1 000 000,头像的百分数更接近 50%:99.7%置信区间从 50%±1.5%缩小为 50%±0.15%。

8. 放入 5 个 0 和 5 个 1。告知抽 1 000 次。

9. 它是好的。一点的次数并不期望恰好是 16.67 次;只预计在那附近。

10. (a)盒子(iii)最佳(猜 0,有 6 分之 5 的机会正确);盒子(i)最差。

    (b)盒子(iii),变化最小;盒子(i),变化最大。

    (c)盒子(iii),SD 最小;盒子(i),最大。

    评论. SD 度量了变异程度。

第 18 章　概率直方图的正态近似

A 组,第 343 页

1. 70 到 80 之间包括端点。

2. (a)7

   (b)7:第二张图里最长的长条。

   (c)不,这只是机会变差,事实上 4 不及 5 象,如最底下一张图中的概率直方图所示。

   (d)(iii),最上面一张是经验直方图——它表示观察百分数,不是概率。

3. (a)3,6

(b)底下一张图—— 概率直方图表示机会。对于积,值 2 和 3 同样可能。

(c)看第二张图:3 更常出现,又是机会变差。

(d)值 14 不可能是积。原因:分解 14 的因子只有两种方式:1×14 或 2×7;无骰子可出示 7 和 14。

4. 底下一张图是概率直方图,故其下面的面积表示概率:11.1%是掷一对骰子出现点数的积为 6 的机会。

5. 不真,和的概率直方图告诉你该和的各种机会。它不告知你抽取的结果如何,阴影面积表示该和在 5 到 10 范围内,包括端点,的机会。(该盒子有 85 张标有 0 的票,2 张标有 1 的票,13 张标有 2 的票。)

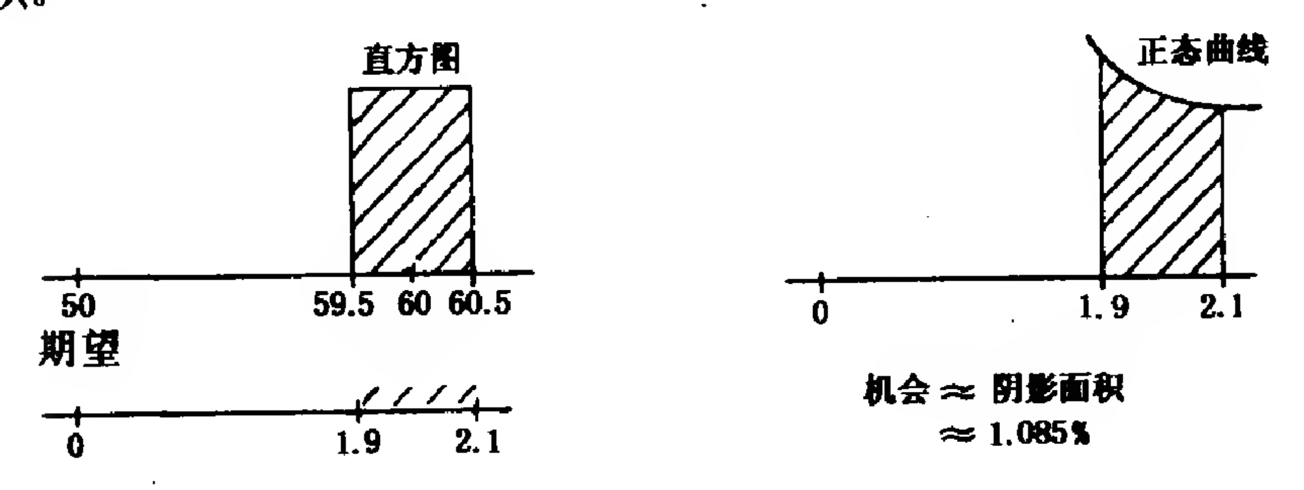
6. (a)根据 A 的和:它的概率直方图更偏右。

(b)根据 B 的和:它的概率直方图较窄,在这个和中有较小的变异。

B 组,第 351 页

1. 直方图下 51.5 和 52.5 之间的面积给出实际机会。正态曲线只是一种近似(但非常好的一种)。

2. 头像的期望次数是 50;SE 是 5,你想求 347 页上图 3 中 60 上面的矩形面积。



评论. 实际机会是 1.084%,在 HP15C 计算器上算得的结果。

3. 根据习题 2,约 100 组中有 1 组应有 60 个头像,事实上 100 组中确有 1 组作到(#6 900——#7 000,这组)

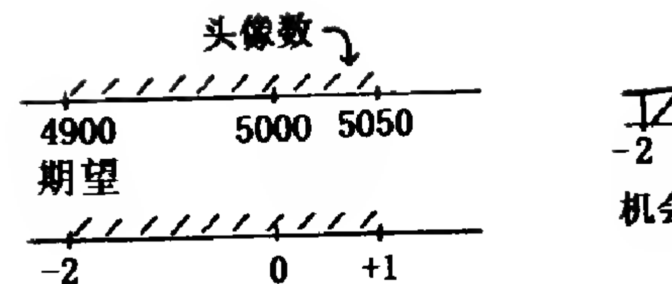
4. 头像的期望次数是 5 000;SE 是 50。(见下页上图)

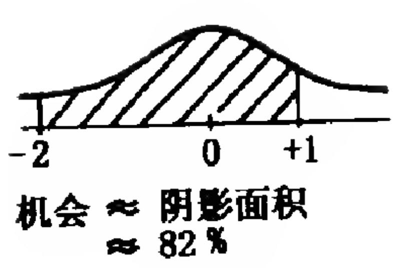
(b)机会≈2%

(c)机会≈16%

5. (a)是,这些块形是大的

(b)不,一些小的块形。





评论. 关于(a)的留意这些边界把估计由 50%变为 54%。

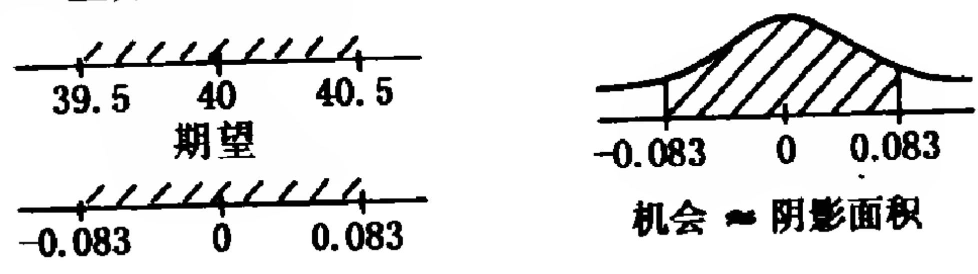
6. (i)正好 6 个头像

(ii)3 到 7 个头像不包括端点

(iii)3 到 7 个头像包括端点。

C 组,第 356 页

1. 掷这枚有偏硬币 400 次中的头像次数等同于取自盒子|9 张[0],[1]|中 400 次抽得数之和。头像的期望次数为 40,SE 为 6。你想求 353 页上图 8 底下一张图中 40 上方矩形的面积。



由图可得这面积在 4%和 8%之间,实际上它是 6.6%,机会亦如此。

2. 在 1 附近正态曲线低于直方图,因此估计将偏低。

3. 是,大的块形。

4. 选(i)。

评论. 机会由概率直方图下的面积给出。经常地,正态曲线下相应的面积是良好的近似,但这里不行——曲线比直方图高得多,因此曲线下的面积高出直方图下的面积许多。

5. 很可能,105;不大可能,101,期望值 100。

评论. 这直方图在期望值近旁有一凹槽。(对于 100 次抽取这凹槽消失。)

6. A(ii),B(i),C(iii)。盒子越不匀称,直方图越偏向一边。

评论. 从盒子|24 张[0],[1]| 中抽取 25 次,你不能期望会得到许多张[1]。概率直方图中最左边的矩形给出抽得数之和为 0(抽得数会为 0)的机会。这个机会是 36%。紧挨着的矩形给出和为 1(抽得数中有一张[1],24

张[0])的机会。这个机会是38%，余类推。(这些机会可用二项公式算出，第15章。)

7. (i)100 (ii)400 (iii)900

当抽取次数增加时，直方图更接近正态曲线。

8. (a)远小于50%。值276 000是0.276百万，在水平轴上约在0.2和0.4的半当中。这点以右的面积远小于50%。这直方图有很长的右尾部，并且期望值比中位数大得多。

(b)1 000 000/100=10 000

(c)400 000到410 000有很大可能：390 000到400 000的机会实际上是0。积有颇反常的概率直方图。

第六部分 抽样

第19章 抽样调查

习题A，第386页

1. 是。即使在八十年代，电话用户可能区别于非电话用户。但是，非电话用户的百分比是如此地小，因此这种偏性通常可忽略，使用电话簿将导致严重偏性，因为有许多号码没被列入，见第7节。

2. 不，你可以预料黑人访问人员询问的答卷人是更多地表示谴责的。(他们是如此。)

3. 不，这个教区可能完全区别于南方的其它教区。(实际上，情况是：Plaquemines是产糖的县，而产糖业比产棉业需要有更高技能的工人。)

4. 总体和样本相同，即，1968年荷兰全体18岁的男人；故无随机误差存在的余地。这项研究将在第612—613页上的习题12中再次讨论。

5. 不，首先ETS关于“代表性”学校的判断可能是有偏的。其次，各所学校可能没有用好的方法抽取它们自己学生的样本。最后，先抽出学校样本再抽学生样本有别于抽学生样本。极易造成过多的学生产生在很小的学校。

评论. 在美国有约3 000所高等学校，其中包括初级大学，社区大学和师范学院。它们中约1 000所非常小，总起来只招收学生总体的10%。另一头，有约100所注册人数都在20 000人以上—它们加起来占学生总体的30%。

6. 选(ii)。

7. 总体由本学期全体登记注册的大学生组成，参数是这些学生中住在家里的百分比。

8. (a)这是一种概率方法：它是完全确定的，机会以有计划方式起作用，至于谁进入样本，无人有权酌情处理。

(b)这方法有别于简单随机抽样。比如，名字在名单上毗连的两人无机会一起进入样本。(简单随机样本在第 4 节中定义。)

9. 颇为不同。不回答者区别于回答者——早回答者区别于迟回答者。(在这项研究中，肺病的百分比在最后 200 人中颇为高些：也许这些人不想进一步证实他们的病。)

10. 样本设计的叙述将比判以不承认的拍卖广告更让人放心。

11. 回答者可能非常有别于不回答者，而仅有样本的 1%回答，相信圣经创世说的教师们可能更渴望回答。

第 20 章　抽样中的机会误差

A 组，第 402 页。

1. 红弹子的数目等同于取自盒子 [1][1][1][1][0] 的 400 个抽得数之和，和的期望值为 $400\times0.8=320$，和的 SE 为 $\sqrt{400}\times\sqrt{0.8\times0.2}=8$。故红弹子数目是 400 中的 320，加减 8 左右。换算成百分数，抽出弹子中红弹子的百分比将为 80%，加减 2%左右，因 400 中的 8 是 2%。

2. (a)盒中应有 100 000 张票子，每张对应总体中的一个人，其中 60 000 张标 1(已婚)，40 000 张标 0，样本中已婚人数等同于取百盒子的 1 600 个抽得数之和，和的期望值是 $1\ 600\times0.6=960$，盒子的 SD 是 $\sqrt{0.6\times0.4}\approx0.5$，和的 SE 是

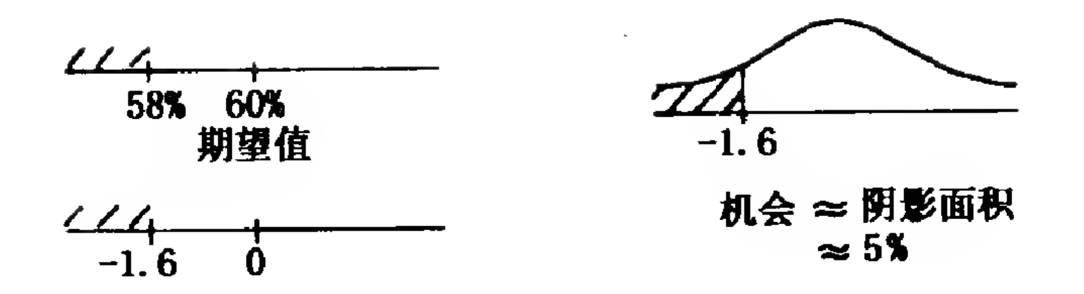
$$\sqrt{1\ 600}\times0.5=20$$

样本中已婚人数将为 960，加减 20 左右，而 1 600 中的 960 是 60%，1 600 中的 20 是 1.25%。故样本中 60%的人已婚，加减 1.25%左右。

(b)盒中应有 100 000 张票子，其中 10 000 张标 1(收入超过 75 000 美元)，其余 90 000 张标 0。有 1 600 个抽得数。机会约为 9%

(c)盒子有 100 000 张票子，其中 20 000 张标 1(大学程度)，其余 80 000 张标 0，有 1 600 个抽得数，机会约为 68%

3. (a)68%左右　　(b)25 的 68%＝17



(c)19 个有 41 到 51 包括端点。

(d)不，这只是机会变差。有约 3%的机会有 5 个 F；3%的机会有 5 个 M；有 20 列：其它规律可能也引人注目，机会变差做了有趣的事。

4. 机会约为 95%

评论。当抽取次数增加时，样本中男人的百分数应更接近总体中的百分数。

5. (a)8%左右，方法参见第 18 章，第 4 节。

(b)50 的 8%＝4　　(c)4

6. 10%±1%。样本中红弹子数是 90±9；若这个数是稍高 1 个 SE，它是 90＋9：将它换算成占 900 的百分数，我们的百分数的 SE 是对期望值加减的量，不是乘以。

7. 不，头像数目的 SE 是$\sqrt{10\ 000}\times 0.5=50$，但在此之后，你应换算成百分数：$(50/10\ 000)\times 100\%=1\%$的 0.5。

8. 不真。他们忘了改变盒子。"1"的个数等同于取自盒子 |0|0|0|1|0| 中的 400 个抽得数之和

9. 选(ii)：机会误差。在总体中假定一个均匀划分(21 以上或 21 以下)，它看来大概是正确的，样本百分数的 SE 约为 2.5%，故(iii)中的误差大了点。

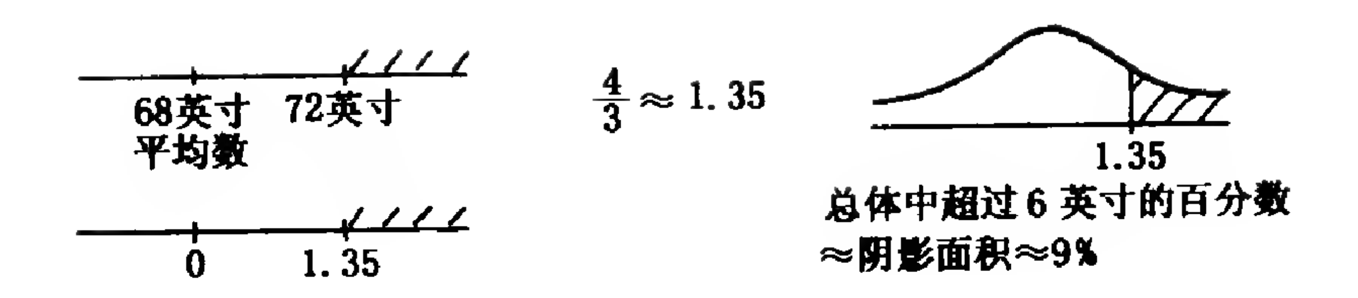
10. 前移的总距离等于掷出点数的总和，这等同于(随机有放回地)取自盒子

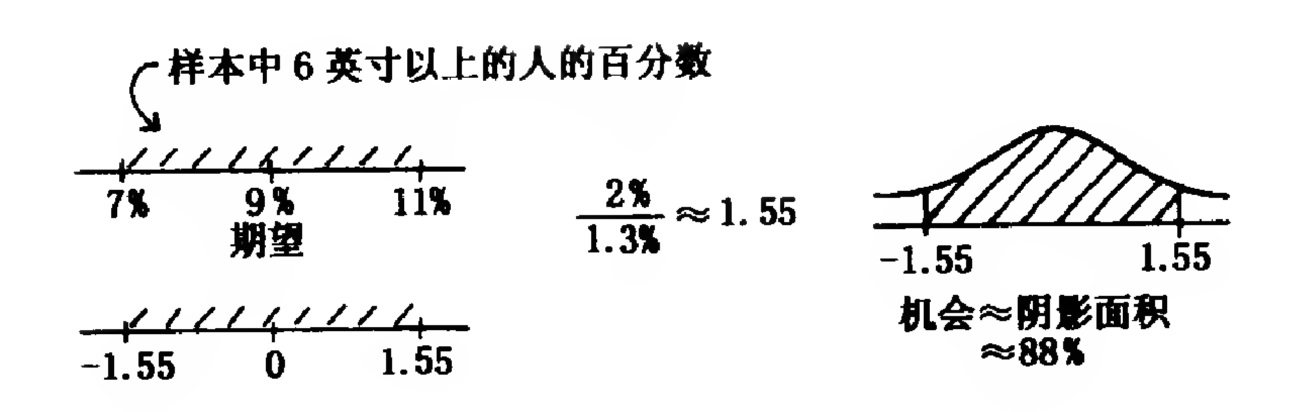
|1|2|3|4|5|6|

中的 200 个抽得数之和，盒子的平均数为 3.5，SD 为 1.7，故他可预期前移约 $200\times 3.5=700$ 个方格，加减$\sqrt{200}\times 1.7\approx 24$ 个方格左右。

11. 对应总体中的每一个人盒子中应有一张票子，对应高 6 英呎以上的人标 1，其余标 0。第一步是计算盒子中"1"的百分数，这可由正态近似来做(第 5 章)：

盒子中"1"占的分数约为 0.09，故盒子的 SD 是$\sqrt{0.09\times 0.91}\approx 0.29$，500 个抽得数中"1"的百分数的 SE 将约为 103%。





B 组，第 408 页

1. 选择(iii)是对的，那是本节的论点

2.

抽取的次数	抽得数中"1"的百分数的 SE
2 500	1%
25 000	1%的 0.27
100 000	0%

评论。抽取 100 000 次后，盒子里已不再有票子，关于抽得数中"1"的百分数不再有不确定性。

3. 样本容量应为 2 500。

4. 有放回时的 SE＝20%，无放回时的 SE＝$\sqrt{\frac{10-4}{10-1}}\times 20\%\approx 16\%$。

评论。这是一个人为的例子，这里抽取的数目占盒子中票子的很大部分，故必须使用修正因子。

第 21 章　百分数的准确性

A 组，第 418 页

1. 317 是取自盒子的 400 个抽得数之和，和的 SE 估计为 8，这就是答案。

2. 第一步是建立模型。盒子里有 100 000 张票子，一部分标 1(对应红弹子)，其余为 0(对应兰弹子)，然后从盒中抽取 1 600 次构成样本。样本中红弹子的数目等同于抽得数元和，盒子中"1"占的比率未知，但可由样本中观察到的"1"占的比率估计，它是 322/1 600≈0.20,。故盒子的 SD 估计为

$\sqrt{0.20\times0.80}=0.40$。样本中"1"的数目的 SE 为$\sqrt{1\ 600}\times0.40=16$。这个 16 是 322 中机会误差的可能大小。"1"的百分数的 SE 为(16/1 600)×100%=1%，盒子中红弹子的百分数估计为 20%，加减 1%左右。

3. 估计值为 39%，加减 2%左右推理见习题 2。
4. 估计值为 48%，加减 5%左右。
5. 估计值为 4%，加减 1%左右。
6. 估计值为 54%，加减 2.5%左右。
7. 不，大多数人为少数几个大公司工作。
8. SE=2%
9. (a)18%±1.9%　(b)21%±2.0%　(c)24.5%±2.2%

   评论。第三个人在估计盒子里"1"的百分数中偏离了两个 SE；即使如此，估计的标准误差仅偏离 1%的 0.2，这证实了自助法有多好。

B 组，第 422 页

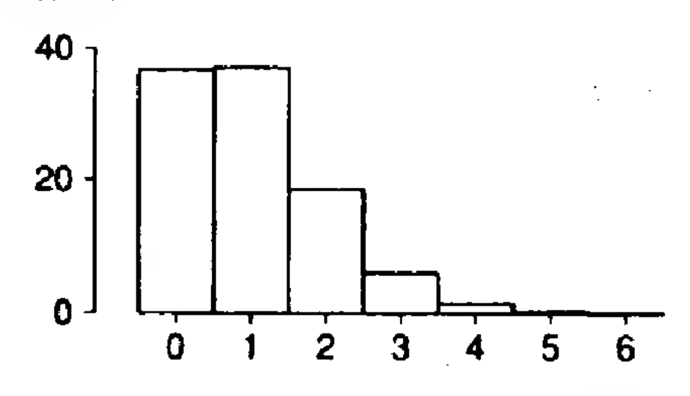
1. 917 是取自盒子的 1 600 个抽得数之和，和的 SE 估计为 20。这就是答案。
2. (a)18%到 22%　(b)17%到 23%
3. 39%±4%
4. 39%±2%

   评论，当样本容量增加时，置信区间变短
5. (a)抽出的弹子中预期有 1 只红弹子，加减 1 只左右。

   (b)不可能抽出少于 0 个红弹子，故机会为 0。

   (c)约 16%。

   (d)不，若概率直方图看起来象正态曲线，则抽取少于 0 个红弹子的机会可由曲线读出。这不行——参见(b)和(c)——故直方图看起来并不象正态曲线，直方图给出如下。



6. 不真。在这里不能用正态近似，我们充其量可根据样本估计，盒子中 1%的

球是红色的，99%是兰色的。这是习题5中的盒子。在从该盒中抽出的100只弹子中红弹子的百分数的概率直方图看起来不象是正态曲线。

C组，第425页

1. (a)18%±3.8%，复盖。

   (b)21%±4.0%，复盖。

   (c)24.5%±4.4%　刚没能复盖。

2. (a)不真。50%是精确地算出的；机会误差是在头像的百分数中，不在期望值中。

   (b)真。

   (c)不真。置信区间是对参数的；在这里，参数——头像的机会——已知为50%

   关于(a)的评论。SE告诉你抛硬币的结果中出现头像百分数的机会误差的可能大小。但是，50%是硬币的特性，它不依赖于抛出的结果，在50%中不存在机会误差。比如，若你抛硬币100次得到53次头像，抛出的结果中出现头像的百分数为53%，机会误差——53%中的——为+3%，若你得到42次头像，抛出的结果出现中头像的百分数为42%，机会误差——42%中的——为-8%。但是期望值50%保持不变，不管抛出的结果如何。

3. (a)真　(b)真　(c)真

   (d)不真：样本百分数为53%，对它你不需要一个置信区间。

4. (a)真

   (b)不真；样本百分数为已知，并在区间中。

   (c)不真：总体百分数要末在区间中要末不在区间中——不涉及机会，参见第423页。

5. 不真。百分数的SE度量了一个样本的百分数与总体的百分数之间差的可能大小，它不度量两个样本百分数之间差的可能大小。

   评论。两个样本百分数间差的SE应该大些，因为它们都蒙受机会变异。相反，总体百分数无变化，更多的有关两个样本百分数间差的内容参见第27章。

6. 真。谈论样本百分数等于这或等于那的机会是可以的——从一个样本到另一个样本百分数是在变化的。在概率的频数理论中，谈论一个总体百分数是这或那的机会是不行的——你考虑的是一个总体，还是几个总体？

D 组，第 427 页

1. 理论说，当心这个人，他在谈论那一个总体？为什么他的学生们是取自总体的简单随机样本？除非他能回答这些问题，否则不必对那些运算多加关注。
2. 这不是简单随机样本，故这公式将不会给出正确答案。

E 组，第 430 页

1. 这不是一个简单随机样本，故公式不适用。
2. 这是好的。
3. (a)机会变差——Gallup 民意测验是基于随机样本

   (b)如表 2 所示，几个百分数点的机会误差是很可能的。毕竟有可能 9 月末并非 11 月初的良好指南。（另一方面，Bush（布什）确实获胜。）

第 22 章　估量就业与失业

A 组，第 446 页

1. 这还行；它是家庭的一个简单随机样本。
2. (a)是：家庭是一个整群。

   (b)不：要估计 SE，你需要更多的信息；第 5 节。
3. (a)不，这不是任何一种概率样本，它是一个方便样本。

   (b)答案与(a)相同。
4. (a)真。

   (b)不真，劳工统计局应该将样本按种族，年龄等等分组，然后对各组分别加权；第 4 节。
5. 111.4 百万±0.1 百万；第 5 节。
6. 百分数的 SE 仅为 1%的 0.2，故差异 68%－55%＝13%几乎不可能解释成一个机会误差，人们喜欢自称他们参加投票，即使他们并没有参加投票。
7. 白人男子的一个；它基于更多的人。

第 23 章　平均数的精度

A 组，第 455 页

1. (a)110/100＝1.1　　(b)0.9×100＝90

2.

抽取的次数	抽得数之和的 SE	抽得数的平均数的 SE
25	10	0.4
100	20	0.2
400	40	0.1

3. 抽得数之和的 SE 为$\sqrt{100}\times 20=200$。平均数的 SE 为 200/100=2，抽得数的平均数将约为 50，加减 2 左右，即使抽取是无放回的，它仍为真，因为盒子中只有很小的一部分票子被抽取。

4. 抽得数的平均数在 2.25 和 2.75 之间的机会，

5. 50 次抽取中 [4] 出现次数的百分数。

6. 平均数的 SE 为 1，故(a)的答案是几乎 100%，(b)的答案是 68%，不要混淆抽得数的平均数的 SE 与盒子的 SD。

7. (a)不真

(b)真，重复一下，不要把抽得数的平均数的 SE 和盒子的 SD 混为一谈。

8. 抽得数的平均数的 SE 为

1，从盒 A　　1.4，从盒 B　　2，从盒 C

(a)203.6 极不可能来自盒 A——它偏离取自盒 A 的 100 个抽得数的平均数的期望值 3.6 个 SE，它也不大可能来自盒 B，因为 3.6/1.4≈2.6 个 SE 也太多，因此，它来自盒 C：类似地有，1981 来自盒 B，由排除法，剩下的 200.4 归盒 A

(b)它可能是另外的，但那将增加一些东西。

B 组：第 462 页

1.

总体，	盒子
总体平均数	盒子的平均数
样本	抽得数
样本平均数	抽得数的平均数
样本容量	抽取个数

2. “SD”一词适用于一个数列，而“SE”适用于机会过程。盒中的票子(和它们的平均数)是固定的，但是抽得数是随机的。

(a)“盒子的 SD”有意义；“盒子的 SE”无意义

(b)“抽得数的平均数的 SE”有意义；“盒子的平均数的 SE”无意义。

3. 用和的 SE，算得它为 540 000 美元，

4. 由样本估计为，样本的 SD 为 4.1 年；这用来估计盒子的 SD

5. 50 的 95%≈48.

6. (a)各机构取它的样本平均数作为它的置信区间的中心。由于机会变差的缘故，样本平均数是不相同的，

(b)样本的 SD 是不相同的(机会变差)，故估计的 SE 是不相同的，这就是为什么区间的长度会不相同。

(c)49

7. 盒子有 30 000 张票子，每张对应一名注册的大学生，标有他或她的年龄，数据等同于取自盒子的 900 个抽得数；样本平均数等同于抽得数的平均数，盒子的 SD 估计为 4.5 岁，抽得数之和的 SE 为$\sqrt{900}\times 4.5=135$ 岁，平均数的 SE 为 $135/900=0.15$ 岁，区间为 22.3 ± 0.3 岁，

8. (a)区间是 368 美元±8 美元，即使数据不遵循正态曲线，抽得数之平均数的概率直方图遵循正态曲线。

(b)不真：8 美元是抽得数的平均数的 SE，不是盒子的 SD。

9. 不真。平均数的 SE 给出样本平均数与总体平均数之间差的可能大小——不是两个样本平均数间的差的可能大小，故 8 美元是一个不该使用的误差幅度。参见第 426 页上的习题 5。

10. 抽得数的平均数只不过是它们的和除以 25(抽取的个数)，故 25 变为 1，50 变为 2，55 变为 $55/25=2.2$，

11. 概率直方图是关于样本平均数的机会的；它不是关于数据的，在这里，概率直方图是给定的，(a)问相对于概率直方图，+1 个标准单位是多少，我们需要这个直方图的中心和散布度，中心是样本平均数的期望值，它等于盒子的平均数。这是给定的：它是 31 700 美元，散布度是样本平均数的 SE，它可以精确算得，因为问题给出了盒子的 SD，这是 20 000 美元，故抽得数之和的 SE 为$\sqrt{400}\times 20\ 000$ 美元，$=400\ 000$ 美元抽得数的平均数的 SE 为 400 000 美元/400=1 000 美元。而+1 标准单位是 31 700 美元+1 000 美元=32 700 美元，这是(a)的答案，

在(b)中，要你找出 30 700 美元在概率直方图轴上的位置它小于期望值：30 700 美元小于 31 700 美元，因此，30,700 美元是在坐标轴的负数部分，事实上，这个值是期望值以下 1 000 美元，而 1 000 美元是一个 SE，故 30 700 美元是－1 个标准单位。这是(b)的答案。

评论,(i)一个有代表性的样本平均数约偏离总体平均数 1 个 SE 左右,我们的样本平均数偏低 1 个 SE 以上,我们取到太少的富人。

(ii)查看 454 页上的图 1,直方图是关于随机抽取过程和取平均数的;它不是针对任一组特定的抽得数,如果你抽 25 张票子而且它们的平均数碰巧为 3.2,这不改变直方图,这道题在更为复杂的背景下描述了相同的论点。

(iii)你应该用值为 20 000 美元的 SD 去换算成相对于该镇全体 25 000 个家庭的收入的数据直方图的标准单位。值为 19 200 美元的 SD 是对另一个相应于 400 个样本家庭的收入的数据直方图起作用。

C 组,第 466 页

1. 95%—置信区间为 1.86±0.09。
2. 这是定性数据,用第 21 章的方法,区间是

   46.8%±7.5%

3. 这不能用正态曲线来做,设样本精确地反映总体,那么公司是从如下的一只盒子中抽样,该盒子里 99.75%是 [1],1%的 0.25 是 [0],这盒子是如此地偏,因此,对于 400 个抽得数之和的概率直方图不可能有一点象正态曲线。参见 422 页上的习题 5,
4. 这是一个整群样本——家庭是一个整群。故 SE 不能用本节的方法估计,半样本方法可替代使用(445 页),但需要更多信息,
5. (a)真,抽得数之平均数的期望值等于盒子的期望值(455 页)

   (b)不能分辨;你需要盒子的 SD
6. (a)不能分辨;你需要盒子的平均数。

   (b)为精确计算 SE,你需要盒子的 SD;即使是估计它,你也需要抽得数的 SD。

   评论,期望值适用于随机抽取过程,不是任一组特定的抽取数,比如,假设你从盒子 [0] [2] [3] [4] [6] 中随机有放回地抽取 25 次,抽得数之平均数的期望值为 3,你的抽得数的平均数可能是 3.1,它高出期望值 0.1;或抽得数的平均数可能是 2.6,它低于期望值 0.4,存在许许多多其他的可能性。但是期望值依赖于盒子,不管抽得数是什么,它保持不变,
7. 盒子的平均数由样本的平均数估计:297/100≈3.0;对 SE,你需要 SD。
8. 这两个过程是相同的:简单随机样本意指随机不放回地抽取(373 页)。

第七部分　机会模型

第24章　测量误差模型

A组，第478页

1. 用和的SE；估计为60毫克，
2. 估计值为测量值的平均数：82，670磅，这可能按平均数的SE（100磅）偏离。
3. (a)800微米　　(b)80微米　　(c)91.4402厘米±160微米。
4. (a)不真。这个范围是从平均数向两边各移2个SE，不是2个SD，
   (b)不真，理由同(a)
   (c)真，
   (d)不真，这恰如第464页上的习题9。
5. 因子是5

B组，第484页

1. 你将不得不抛该图钉多次，并观察它落下时尖头向下次数的百分数是接近于50%还是67%。（这将依赖于表面：在一次试验中，当在漆布上抛时，图钉尖向下占66%次，，但在地毯上抛时，只占50%次。）
2. 不，在雨季，雨天全接近一起出现，若某天下雨，则很可能第二天也下雨。
3. 最后的数字，是，第一个数字，不是，比如，在San Francisco（旧金山）的电话簿中，第一个数字不可能是0。另外，比起2来有更多电话号码是9字当头。
4. 不，字母按顺序出现，无一盒子将适合它。
5. (a)抛一枚硬币是数字的机会理论可直接用于它的一种情况，进行一项体育运动则不是，如果你试图提出体育运动的机会模型，使相继的比赛相互独立是错误的。
   (b)象投篮，你有50对50的对半分机会赢5美元或输4美元

C组，第488页

1. 在两种情况中，测量值均是10克之上504毫克。
2. 不，如前一道题所示。
3. 6毫克是111页上表1所录100个测量值的SD，它被用来估计误差盒子的

SD，因此，“由数据估计为”

4.(a)机会变差——调研人员得到不同的样本平均数。

(b)又是机会变差——调研人员得到不同的样本 SD。

(c)50 个区间中约 95%，即约 48 个区间，将复盖确切重量。

(d)48(其中有一个区间有较大偏差，——机会变差在起作用)

5. 误差盒子的 SD 估计为 50 毫克。

(a)5 毫克——平均数的 SE。

(b)50 毫克——误差盒子的估计 SD。

(c)95%——两个 SE

6. 答案是 1.2 毫克，参见例 5

7.(a)300 007(平均数)，2(平均数的 SE)，

(b)真：数列中的每一个数偏离该数列的平均数一个 SD 左右。

(c)不真：平均数精确为 300 007，

(d)不真：2 是 SE，不是 SD

(e)真：区间是“平均数±2 个 SE”。

(f)不真，25 个测量值的平均数精确为 300 007。

8. 答案是 2 英寸。这里给出理由，每一测量值等于精确长度加上取自误差盒子的一个抽得数。估计距离 AE 是 4 个测量值的和，并偏离 AE 的精确长度约 4 个取自盒子的抽得数之和。误差盒的平均数为 0，故 4 个抽得数之和将约为 0，加减一个 SE 左右。和的 SE 是正确的加减的量，盒子的 SD 是 1 英吋，故和的 SE 为$\sqrt{4}\times 1$ 英吋＝2 英吋，

评论. 求 AE 的长度涉及测量 和，不是求它们的平均数。

9. 不同人的机会误差可能有不同的 SD，此外，若同一人做检测若干次，误差可能相依。Gauss 模型似乎不适用。

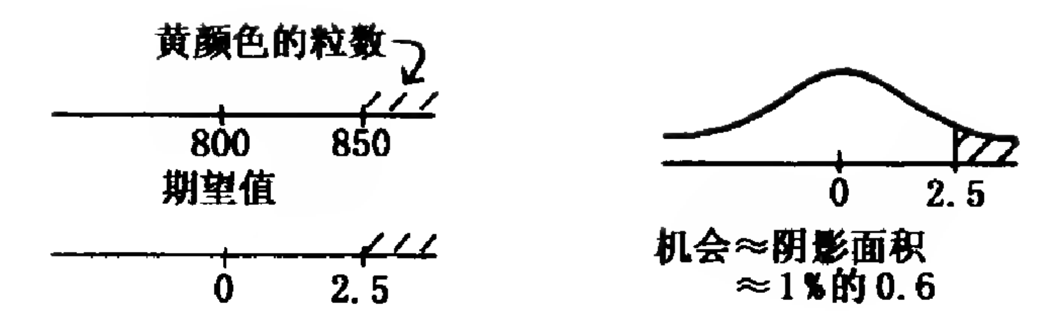
第 25 章　遗传学中的机会模型

A 组，第 499 页

1. 每粒种子有 50%的机会由 y/g 亲本中获得 y，以及 505%机会获得 g，它从 g/g 亲本中肯定获得 g，故该种子有 50%机会为 y/g 颜色是黄的；它有 50%机会为 g/g，颜色是绿的约 50%的种子应为黄色。

1 600 粒种子中黄颜色的粒数等同于取自盒子 [0][1] 的 1 600 个抽得数之和，黄颜色的期望粒数为 1 600×1/2＝800，SD 为$\sqrt{1\ 600}\times 1/2=20$。现可用

正态近似。



2.(a)白×红→100%粉红

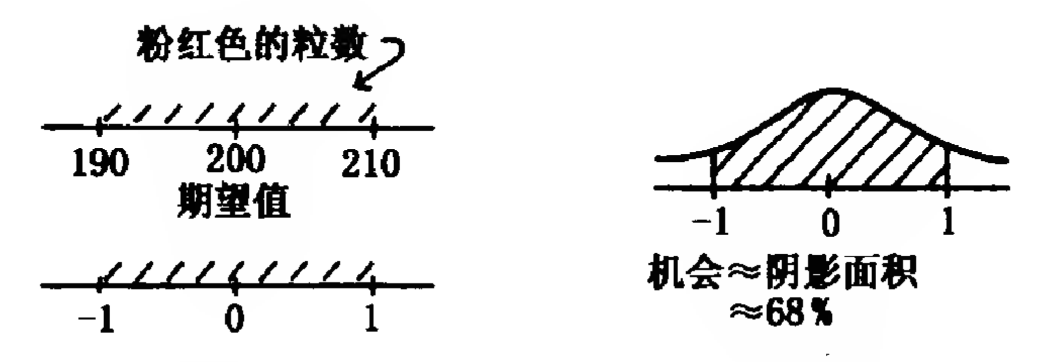
白×粉红→50%白,50%粉红

粉红×粉红→25%红,50%粉红,25%白

求粉红×粉红:每一亲本为r/w,故后代花朵颜色依据从下表中随机取一行和一列决定

	红	白
红	红	粉红
白	粉红	白

(b)400颗作物中预期粉红色的数目是200,附一个值为10的SE,用正态近似。



3.(a)某基因对依据变异体w宽(和n(窄)控制叶子的宽度。规律为:w/w趋于宽,w/n和n/w趋于中等,n/n趋于窄。

(b)窄×窄=n/n×n/n→100%n/n=窄

窄×中=n/n×n/w→50%n/n窄,50%n/w=中等

4.B=棕色,b=蓝色,丈夫是B/b,妻子是b/b,每个小孩有1/2的机会是棕色眼睛,三个小孩相互独立,故三个小孩都是棕色眼睛的机会是$(1/2)^3=1/8$。

第八部分　显著性检验

第26章　显著性检验

A组,第518页。

1. (a)3 292美元−3 117美元=175美元。

   (b)在新条例下,她的处境较好,因为她付较少的税。

2. (a)100 000张在盒子中并抽100张。

   (b)由数据估计为

3. 两套条例下的平均数之间的差是3 182美元−3 217美元=−35美元。若新条例是税收一中性的,期望差为0,35美元约离这期望值半个SE:

$$\frac{-35\text{美元}-0\text{美元}}{72\text{美元}}\approx-0.5$$

   助手获胜:这个差看似机会变差。

4. 如果骰子是公平的,点数的总和等同于取自盒子 [1][2][3][4][5][6] 中的100个抽得数之和。盒子的平均数是3.5,SD是1.7。故和的平均数是350,SE是17。点数稍大于它的期望值1个SE,它看似机会变差。

5. 问题可以象习题4那样提出,但是这一次点数约大于它的期望值3个SE。这看来不象是机会变差。

   评论.(i)样本容量有关系;比较一下习题4和5。

   (ii)关于骰子公平性的更完整的检验将在第28章中给出。

B组,第520页

1. 不真,那是原假设。
2. 真
3. 一个数据的模型。原假设是有关盒子模型的一个陈述,而检验告诉你这个陈述是否似乎可能的。
4. 盒子的SD可以估计为10,故100个抽得数的平均数的SE可估计为1。若盒子的平均数是100,则抽得数的平均数大于它的数望值2.7个SE。这似乎不是可能的。若抽得数的平均数是101.1,那是在机会变差范围之内——1.1个SE。

C 组，第 523 页

1. (a)77%　　　　(b)1%的 0.1

   大的 P 有利于原假设，小的 P 不利于原假设。

2. (a)真；参见 521 页。

   (b)不真；参见 523 页。

3. 平均数的 SE≈1.25，故 Z≈(52.7−50)/1.25≈2.2，而 P 近似为 2.2 以右正态曲线下面的面积。由表，这约为 1.4%。这个差难于解释为机会变差。

4. 样本等同于随机取自上下盒子的 100 个抽得数，该盒子对应每名雇员有一张票表明缺席的天数。原假设：盒子的平均数是 6.3 天。备择假设：盒子的平均数小于 6.3 天。盒子的 SD 估计为 2.9 天，故平均数的 SE 为 0.29 天，Z≈(5.5−6.3)/0.29≈−2.8，故 P≈1%的 0.3。这是拒绝原假设的有力证据。

5. 现在 Z≈(5.9−6.3)/0.29≈−1.4，故 P≈8%。原假设看来似乎是可能的。

D 组，第 525 页

1. 盒子，零假设是关于盒子的描述。

2. (a)不真. 即使零假设为真，有 1%的时候实验将给出“高度显著的”结果。

   (b)不真；523 页。

3. (a)真，大的 P 值有利于原假设。

   (b)真，小的 P 值不利于原假设。

4. (a)真；524 页

   (b)不真；P 必须小于 1%。

   (c)真；524 页

   (d)真；522 页

   (e)真；Z＝(观察的值－期望的值)/SE，其中“期望的值是在原假设基础上算出。

5. 约 2%

6. (a)真；Z＝(观察的值－期望的值)/SE；“观察的值”指抽得数的平均数；“期望的值”指盒子的平均数，给定为 50。

   (b)约 50

   (c)他们中约两三个解。

   (d)约 2%。参见习题 5。

E 组，第 529 页。

1. 选择(i)："等同于"意指"就机会涉及的范围之内"。每个猜测有 4 分之 1 的机会正确，而每次抽取有 4 分之 1 机会为 1。于是正确猜测的次数等同于抽得数之和，而且平方根法则适用。

   选择(ii)是错的；抽得数的平均数包含一些机会误差，盒子的平均数则不包含。选择(iii)无助于 SE。选择(iv)是愚蠢的。

2. (a)抛硬币等同于从 0—1 盒子随机有放回地抽取 10 000 次，这里 0＝背面，1＝头像。盒中 1 占的比是未知的。原假设：这个比是 1/2。备择假设：它大于 1/2。头像的次数等同于抽得数之和。

   (b)$Z \approx 3.35$，$P \approx 4/10\ 000$。

   (c)有太多的头像以致不能解释为机会变差。

3. (a)与 1(a)相同

   (b)$Z \approx 1.35$，$P \approx 9\%$

   (c)硬币似乎是公平的

4. (a)那学期在 Berkeley 注册的学生；盒子是关于总体的。

   (b)1＝男人，0＝女人(如果你是对男人计数的话)。

   (c)盒子里有 25 000 张票子，并有 100 个抽得数；样本等同于抽得数。

   (d)100 个抽得数

   (e)67%

   (f)由原假设给定。该盒子是关于总体的，并给你那学期男生的百分数是 67%。

5. (a)53　　　　(b)67

   (c)由原假设计算：100×0.67。期望常由原假设计算。

   (d)和

   (e)$\sqrt{100} \times \sqrt{0.67 \times 0.33} \approx 4.7$

   (f)由原假设计算(它告诉你盒子的组成成份)。

   (g)$Z \approx (53-67)/4.7 \approx -3$，和 $P \approx 1/1\ 000$。

6. 不. 他获得太多的妇女；而 P 太小，故机会将解释不了差。随意地取人与简单随机抽样不同(第 19 章)。

7. (a)原假设：猜对的次数等同于取自有 1 张票为 1 与 9 张票为 0 的盒子中 1 000 个抽得数。

   (b)$\sqrt{0.1 \times 0.9}$. 原假设告诉你盒子里是什么。使用这点。

(c)$Z\approx(173-100)/9.5\approx7.7$,P 是极小的。

(d)无论怎样,它不是机会变差。

8. 数据由 25 个重量组成。原假设称数据等同于随机取自某盒子的 25 个抽得数。对集群里的每只动物,盒子里有一张票子标照它的重量。故盒子的平均数是 30 克。SD 是 5 克,故 25 个抽得数的平均数的 SE 是 1 克。于是 $Z=(33-30)/1=3$,$P\approx1/1,000$。

   评论:(i)这里,原假设告诉你盒子的 SD,因此你不需要由数据去估计它:解题中没有用数据的 SD。

   (ii)随意地选取不同于抽取简单随机样本(第 19 章)。当你伸手进笼子里捡出一只动物时,可能你抓在手里的是较驯服的,它们比其它的略重些。

9. 原假设称降低价格对销售量没有影响。故在每一对商店里,按通常价格的一家和它的按降低价格的配对差不多售出的一样多。按照盒子模型,原假设称数据等同于取自盒 | 1 | 0 | 中的 25 个抽得数。这里 1 表示售出的较多的正常价格商店,0 表示售出较多的削价商店。1 的期望次数是 12.5,SE 是 2.5,故 $Z=(18-12.5)/2.5=2.2$,$P=1.4\%$。拒绝原假设的证据相当强烈。

   评论. 这方法称为"符号检验"。

F 组. 第 538 页

1. (a)5%　(b)5%　(c)90%　(d)95%
2. 由表,2.92 以右的面积为 5%,6.96 以右的面积为 1%。由于 4.02 界于 2.92 和 6.96 之间,故 4.02 以右的面积界于 1%和 5%之间。
3. 不,3 个自由度。
4. (a)自由度=2,平均数≈72.7,$SD^{+}\approx5.7$,SE≈3.3,$t\approx(72.7-70)/3.3\approx0.8$

   P 约为 2.5%。推断:校正是好的。

   (b)P 约为 2.5%;重校。

   (c)一次测重是决不够的。

   (d)P 约为 25%

   关于(d)的评论。两次测量优于一次测量,但越多越好。
5. 在(a)中,数 93 是一个异常值,故误差似乎不遵循正态曲线。在(c)中,数目

只是在 69 和 71 之间来回转换。这不利于 Gauss(高斯)模型。

6. 根据 Gauss 模型,10 个新测量的重量,每一个等于确切重量,加上偏度,加上取自误差盒子的一个抽得数。原假设称偏度为 0;备择假设称存在某一偏度。误差盒子的 SD 估计为$\sqrt{10/9}\times 9\approx 9.5$。(误差属于再生尺度,故与 7 毫克的原来的 SD 无关。)平均数的 SE 为 3 毫克,$t\approx -2.67$。-2.67 以左 9 个自由度学生氏曲线下面的面积约为 1%。有强力证据拒绝原假设。

7. (a)根据 Gauss 模型,100 个测量中的每一个等于确切重量,加上取自误差盒子的一个抽得数。误差盒子里的票平均数为 0。未知参数是确切重量。原假设称这仍比 1 千克高出 512 毫克,备择假设称确切重量较小。

   (b)误差盒子的 SD 可以根据以往数据估计为 50 毫克——误差盒子属于设备。(52 毫克的新 SD 是无关的)。

   (c)100 个测量值的平均数的 SE 为 5 毫克,故

   $$2=(508-512)/5=-0.8,$$

   $$P\approx 21\%$$

   (d)重量的下降看似机会变差。

第 27 章　再论平均数检验

A 组,第 547 页。

1. 期望值为 100−50=50,SE 为$\sqrt{2^2+3^2}\approx 3.6$。由于抽得数全都独立,平方根法则适用。

2. 各百分数的期望值为 50%;诸 SE 是 2.5 和 5 个百分数点。差的期望值是 0,SE 是$\sqrt{2.5^2+5^2}\approx 5.6$ 百分数点。由于两个百分数是独立的,平方根法则适用。

3. 真.若抽取是有放回的,两个平均数将相互独立,并且差的 SE 将确切地等于$\sqrt{3^2+5^2}$。盒子是如此之大,因此在有放回与不放回抽取之间无实际差别。

4. 不真.盒子里只有 200 张票子,两个平均数完全相依——当你知道一个平均数时,你自动地知道另一个。平方根法则将不适用。

5. 盒子 A 的 SD 可以估计为 3,故取自盒子 A 的 100 个抽得数的平均数的 SE 为 0.3;类似地,取自盒子 B 的 400 个抽得数的平均数的 SE 可估计为 0.4;平均数是独立的,故差的 SE 为$\sqrt{(0.3)^2+(0.4)^2}=0.5$。若这两只盒子有相同的平均数,观察差 51−48=3 与期望值 0 的偏差达到 6 个 SE。看

来不可信。

B 组. 第 552 页

1. 两样本 Z—检验.

2. 数据由男士的 1,600 个 0 和 1(1=文盲),和女士的另外 1 600 个 0 和 1 构成。模型有两只盒子,M 和 F。盒子 M 对应县内每位男性青年有一张票,文盲标 1,有文化的标 0。盒子 F 类似地对应女士。男士的数据等同于取自盒子 M 的 1,600 个抽得数,对女士也一样。原假设:两只盒子里 1 的百分数相同。备择假设:盒子 M 里 1 的百分数较大。男士样本里 1 的百分数的 SE 可估计为 1%的 0.6;对女士样本,SE 是 1%的 0.4。故差的 SE 为 $\sqrt{(0.6)^2+(0.4)^2}\approx$1%的 0.7。于是 $Z\approx(7-3)/0.7\approx5.7$,P 几乎为 0。这个差几乎不可能解释为机会变差。

3. 两个平均数的差的 SE 可估计为

$$\sqrt{(0.5)^2+(0.5)^2}\approx0.7$$

故 $Z=(26-25)/0.7\approx1.4$,$P\approx8\%$。这个差可较好地归因于机会。

4. $Z=1/0.45\approx2.2$,$P\approx1.4\%$。

评论. 观察的显著水平依赖于样本容量。对于大样本,即使小的差有可能是高度统计量显著的。在第 29 章中将有更多这方面的叙述。

5. 处理组与对照组的平均数是相依的,因为配对里的小鼠取自同一胎。因此,若一只小鼠有重的大脑灰质,配对里的另一只大概也如此。SE 的计算没有考虑到这一配对关系。

评论. 更好的分析可参见第 26 章的复习题 12。

6. (a)各样本百分数的 SE 为 1.6%,两个百分数间差的 SE 为 2.3%。两个样本百分数间的差为 3.4%,故 $Z\approx1.5$,$P\approx7\%$。这可能是机会。

(b)1974 年的百分数的 SE≈0.7%,1984 年的 SE=1%,差的 SE 是 1.2%。样本百分数间的差是 4.6%,故 Z 是 3.8,P 小于 1/10,000。这是拒绝原假设——服用可卡因的百分数上升——的有力证据。

7. 两个样本平均数间的差为 3 小时,SE 为 0.5 小时。故 $Z\approx6$,$P\approx0$。这个差很难解释成机会变差。私立大学的学生来自较富裕的家庭,有来自家里的较多赞助。

8. 分子是百分数,分母是一个小数。实际上,$Z=(53-48)/5.3$;或 $Z=(0.53-0.48)/0.053$。

评论. 忘了将分母换算成百分数是一个常见的疏忽。你可以按百分数或小数解整个问题;不可半途换马。

C 组,第 558 页

1. (a)数 A 表示一个学生,如果给他(或她)给予辅导时,其 SAT 成绩将是多少;数 B 表示没有辅导时的成绩。

   (b)若一个学生是指派到辅导组,则观察到的是数 A,若是指派到对照组,则观察到的是数 B。你绝不可能两张票子都得到,因为没有一个学生同时指派到处理组和对照组。

   (c)取一条保守的途径:辅导组平均数的 SE 是 9.8 分;对照组平均数是 10.3 分;对于差$\sqrt{9.8^2+10.3^2}\approx 14.2$ 分。平均得分中的差是 9,故 $Z\approx 9/14.2\approx 0.63$,$P\approx 27\%$。这可能无疑是机会变差。

   评论. 这问题与(第 2 节)比较 1982 年与 1973 年中的 NAEP 考试成绩不同,因为我们没有两个独立的样本。但它与维生素 C 试验(例 4)相同。200 名学生的每一人有两种可能响应—— 一个是若被辅导的,一个是若不被辅导的。调研人员只能看到两种响应中的一个,并随机做出他们的选择。那就是为什么 SE 的计算是合理的原因。(554—557 页。)

2. (a)差是 66－59＝7 分,SE 是 1.8 分。故 $Z\approx 3.9$,$P\approx 0$。这个差难于解释为机会变差。Wheafies 起作用。

   (b)学生们将知道他们正在食用的是那一种谷类食物,因此,那方面的研交难于采取盲式。最后的评分可按盲式进行。应该在随机化之前就获知同意参加这项研究,不是在随机化之后,以避免选择缺失。

3. (a)差是 1 分,SE 是 1.7 分。这看似机会变差。两个组是可比的—— 随机化起作用。

   (b)现在差是 9 分,具有相同的 SE,1.7 分,故 $Z\approx 5$,$P\approx 0$。随机化出了差错。

   评论. (b)中的差不能解释为吃 Whiefies 的结果,因为直到期中之后谷类食物供给仍未开始。参见第 23 页上的习题 7。

4. (a)两个样本平均数间的差是 0.1,SE 是 0.13,故 $Z=0.8$,$P=21\%$。这看似机会变差。

   (b)存在一组新的随机数,并且其它因素可能也在起作用—— 天气,新的感冒病毒,等等。毕竟这些研究涉及两个不同时期的两组不同的人。

D 组. 第 562 页。

1. (a) | 0|1 |

   (b)形式 A,赞成手术;形式 B,赞成射线疗法。

   (c)仅(ii)

   (d)得到形式 A 的学生人数是 84+112=196;这些学生中 112/196×100% =57%赞成手术。得到形式 B 的学生中约 83%赞成手术。百分数的差是 26%,SE 约为 5.2%,故 $Z\approx5$,$P\approx0$。这个差难于解释为机会变差。

2. "百分数"意思是"每 100 里的",但是这问题的比率是如此地小,因此将它们表成 100 000 分之几更方便。在疫苗组的比率是 57/200 000 或每 100 000 中的 28.5。病例数的 SE 是

$$\sqrt{200\ 000}\times\sqrt{\frac{57}{200\ 000}\times(1-\frac{57}{200\ 000})}\approx8$$

(捷径方法参见第 17 章第 4 节)。故比率为 SE 是 8/200 000 或每 100 000 中的 4。在安慰剂组中,比率是每 100 000 中的 71,SE 是每 100 000 中的 6。差的 SE 是

$$\sqrt{4^2+6^2}\approx100\ 000 \text{ 分之 } 7$$

比率中的差是 28.5−71=每 100 000 中(−42.5)。在原假设下,比率中的期望差为 0。故 $Z\approx-42.5/7\approx-6$。比率中的差不能解释为随机化中的侥幸。疫苗起了作用。

3. (a)两个样本平均数间的差是 900 小时,SE 约为 300 小时。故 $Z\approx3$,P 约为 1 000 分之 1。差难于解释为机会变差。结论:负所得税使人们少工作,但是影响不大——3 年期间 900±300 小时。(3 年期间 900 小时约为每周 6 小时。)

   (b)两个样本百分数间的差是 6%,SE 约为 3%。这个差难于解释为机会变差。

   评论. (i)负所得税使人们工作得少些;对工作的主妇的影响似乎最大。

   (ii)计算做了一个含蓄的假设,即家庭对负所得税的响应不依赖于给其它家庭的税收待遇。如果这一假设是严重错误的话,则负所得税的影响难于在样本基础上进行研究。

4. 这问题不能由所给信息回答。调研人员设有两个独立样本,一个样本回答关于大不列颠的问题,另一个回答关于法国的问题,故例 3 的方法不适用。调研人员只有一个样本,且样本里的每一个学生有两个响应:

1 1　在地图上找到大不列颠和法国。

1 0　找到大不列颠;不能找到法国。

0 1　不能找到大不列颠,找到法国。

0 0　两个国家都不能找到。

调研人员给考试打分时,两种响应都观察到;这使得它与第4节里的实验不符。(如果你知道以上所列4种类型中每一种的百分数,这问题可以用更高等的统计方法解答。)



5. (a)两个样本百分数之间的差为0.6%,SE约为3.6%。这看似机会变差。劝阻补助进食对稍后的母乳喂养没有影响。

(b)差是20.9毫升/日,SE是3.1/毫升/日。这几乎不可能解释为机会变差。喂养方式看来受到保育室里不同处理的影响。

(c)差是0.9%,SE是0.14%。故Z=6.4。劝阻补助进食增大体重下降:一个不利的付作用。

(d)两个样本平均数之间的差是27克,SE约为31克。这看似机会变差——随机化是成功的。

评论.(i)(c)中存在微妙之处。每个婴儿的体重下降是按相对于初生体重的百分比测量。这些百分比是定量数据,并对它们进行平均数和SD的计算。

(ii)实验表明劝阻补助进食没有促进母乳喂养,并有不利副作用——体重下降。观察研究误解了它。解释:存在一个重要的混杂变量。有教养的母亲们在医院里更可能以母乳喂养,她们的婴儿吃少量补助食品。这些母亲稍后也更有可能是母乳喂养的。因此,在医院里人工喂养与稍后母乳喂养之

间存在负关联。但是这一关联受第三个因素——母亲的品格个性——驱使。

第 28 章　χ^2-检验。

A 组，第 577 页。

1. (a)90%　(b)10%　(c)1%

2. 约 10%

   评论. 将它与 1(c)比较。当自由度上升时，曲线偏向右并伸展开，因此，具 9 个自由度 15.09 以右的面积比具 5 个自由度的大。见图 1。

3. $\chi^2=13.2$，d=5，1%<P<5%；实际上 P≈2.2%。

   评论. d=自由度。数据对模型的拟合并不那么好。

4. $\chi^2=1$，d=5，P≈95%。

5. $\chi^2=10$，d=5，5%<P<10%；实际上，P=7.5%。

   评论. (i)数据对模型的拟合不是特别好。

   (ii)比较习题 4 和 5。观察频数仅乘以 10 倍；这不改变百分比。但是 χ^2—检验的结果依赖于样本容量。对大样本，χ^2—检验将拒绝非常合理的模型。在下面的习题和第 29 章里有更多关于这方面的情况。

6. $\chi^2\approx18.6$，d=5，p<1%——虽然对大多数目的来说，骰子可能与所希望的那样公平；第 29 章里有更多关于这方面的情况。

7.

	状态	机会	期望的	观察的
偶数，大	4，6	2/6	200	183
偶数，小	2	1/6	100	113
奇数，大	5	1/6	100	88
奇数，小	1，3	2/6	200	216
$\chi^2\approx6$，	d=3	P≈10%		

8.

年龄	观察的	期望的
21 到 40	5	27.7
41 到 50	9	15.2
51 到 60	19	10.6
60 和以上	33	12.5

$\chi^2\approx61$，d=3，p≈0。对简单随机抽样来说，一个大陪审团与郡的年龄分布差别得如此之大几乎不大可能。推断是大陪审团的成员不是随机选取的。

评论. (i)期望频数不必是整数。

(ii)大陪审团成员由法官任命,法官们更喜欢年龄大一些的陪审员。

9. (a)12.

(b)非常有趣的事:29 大于机会水平许多个 SE。

(c)你应该查看所有号码的数据,不仅限于#7,并使用 χ^2—检验。χ^2—统计量:A)15.2,B)26.7,C)7.5,D)16.5。对于 9 个自由度,10%水平是 14.68,5%水平是 16.92,1%水平是 21.67。故 A 是边缘状态,B 是界线外状态,C 是好的,D 是边缘状态。

(d)关于重新检验,组 A 的 χ^2 是 14.5,D 的是 18.8。拒绝 D,可能也拒绝 A。

10. 这不是一个好方法。χ^2 的公式涉及频数——数;不是百分数。比较上面的习题 4 和 5。

B 组,第 582 页。

1. 合并的 $\chi^2=13.2+10=23.2$,$d=5+5=10$,$P\approx1\%$。
2. 不,相依的实验。
3. $\chi^2\approx0.5$,$d=3$,$P\approx8\%$。不明确,但针对捏造。

C 组. 第 587 页。

1.

观察的			期望的		差	
2 792	3 591	6 383	2 730.6	3 652.4	61.4	−61.4
1 486	2 131	3 617	1 547.4	2 069.6	−61.4	61.4
4 278	5 722	10 000				

$$\chi^2\approx6.7, d=1, P\approx1\%$$

评论. (i)65%的男人参加投票对比 63%的妇女参加投票。这是一个小的差,但对大样本来说它是精确估计。所有的 P 告诉你的是差是否可以用机会解释。在第 29 章中将有更多这方面的情况。

(ii)一张 2×2 的表,可以用 X^2—检验或 Z—检验(第 27 章,第 4 节)处理;但两种检验将给出略有差别的 P—值:参见第 27 章的注 3。

2.

观察的			期望的		差	
21	9	30	14.0	16.0	7.0	−7.0
20	39	59	27.5	31.5	−7.5	7.5
7	7	14	6.5	7.5	0.5	−0.5
48	55	103				

$$\chi^2\approx10, d=2, p<1\%。$$

一般地，妇女比男人较早结婚；在年龄组 25—29 岁，有比期望的更多的妇女结婚。（“期望的”意指，根据原假设男人和妇女有相同的婚姻状态分布。）超出数额的丈夫在高年龄组，比如 30—34 岁。

第 29 章　显著性检验的进一步考虑

A 组，第 593 页。

1. (a)真，　(b)真。
2. (a)不真，参见第 523 页。　(b)不真。

B 组，第 597 页。

1. (a)约 5　(b)4　(c)约 1

   评论. 若你抛 100 枚硬币，你预期可得约 50 个头像。若原假设为真，得到一个“显著的”结果的机会是 5%；故你可预期它在 100 次中出现 5 次。

2. (a)25　(b) | 0 | 0 | 0 | 1 |

   (c)序号之和等同于随机有放回地取自盒子 | 1 | 2 | 3 | 4 | 的 25 个抽得数之和。

3. (a)真. 在原假设下，高等级的次数等同于取自盒子 | 0 | 0 | 0 | 1 | 的 25 个抽得数之和，因此，一个“显著的”结果的机会约为 3%。

   (b)真　(c)真

4. 略大于。第一种检验单独地有 3%的机会得到一个“显著的”结果；第二和第三种各有约 5%的机会。但三种检验中至少一个发现“显著的”结果的机会大于 5%。

   评论. 对数据窥探的烦恼是它使得显著水平接近于无意义。那些窥探数据的人，即使在没什么事发生下也会有所发现。

5. 又是数据窥探。若 25 个不同的假设都被检验，某些结果可能是显著的。这看起来不象是好的证据。
6. 双尾的。
7. 单尾的。
8. (a)是；　p≈4%。

   (b)不；　p≈96%。

   (c)不；　p≈8%。

   评论. 在接近的情况，使用单尾检验可以使你超越显著线。

9. 医生们如果有不寻常的高死亡事故率时，他们更有可能给刊物写文章，而

那很可能是依据小样本的——它给侥幸留有过多的余地。就象 chalmers 说的，“内科医生们有一种报导不平常事物的倾向”。

C 组，第 602 页。

1. (a)不真　　(b)不真. 参见第 601 页。

2. 问题有意义，因为我们处理的是简单随机样本，它可以用两样本 Z—检验解答：

　　男人的平均数的 SE≈1，妇女的平均数的 SE≈1

　　差的 SE≈$\sqrt{1^2+1^2}$=1.4，Z≈1.4，P≈8%(单尾)

　这可能是机会变差。

3. 各平均数的 SE 为 0.5，故差的 SE 为 0.7，Z=2.8，P 降为 1%的 1/4。

　评论. 观察到的显著水平依赖于样本容量。对于较小的样本，差估计为 2±1.4 分；对于较大的样本，2±0.7 分。

4. 拒绝一个原假设可以依据一个不显眼的差——若样本大的话。但是，对于小样本，一个相当大的差可能够不上统计显著。(596 页.)

5. P=97%. 大 P 对原假设有利，小 P 则不利。

6. (a)要断定按时返回的评审意见更为否定的程度有多大，你需要平均数和 SD，不只是 P。

　(b)71.5%与 25%之间的差是令人吃惊的，并告诉你许多有关评审过程的东西。要想知道 P—值又添了些什么，设想刊物编委们的话，

　　看，我们的评审员中某些人比其它人更为苛刻。由于抽取中的侥幸，太多苛刻的人被选中拿到否定描述形式。

　　P—值告诉你，编委们不能板着面孔用“抽取中的运气”这一防御手段，但是 P—值不帮你比较 71.4%和 25%。

　评论. 在(a)中有另一问题。虽然数据作为试验的一部分被收集，但是这种比较是观察的。对于评审员只是将他们选出，并没有指派他们到“误期”组或“按时”组。参见第 2 章第 2 节。这种评审的研究，在结果的提出中有少量失误，确实有趣。

7. P—值不度量差的大小，因此不可能只根据 P 分辨影响是强或弱。

8. 99%—置信区间是−6.0±2.6 个 SE，即−6±6.5。这个估计不很精确。该 P 值指出顺应性不是正好为 0；没人说它是。

D 组，第 605 页。

1. 在研究中，把旷工情况与前几年的相比，但是这一年可能与上一年不同。

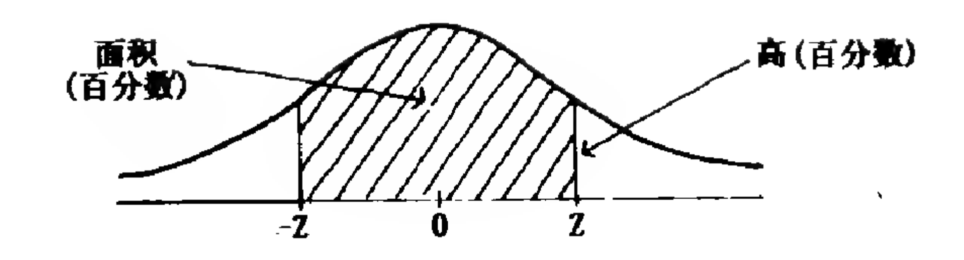
(天气比较暖和，工作更有趣，等等。)最好把弹性工作时间下工人中的旷工总数与同期对照组中的相比较。为了避免那些没给他们弹性工作时间的人的抱怨和不满，对处理或对照规定全部工作单元是个好主意。

2. 这个实验是精心设计的。下结论说疫苗防止了儿童患小儿麻痹症是公正的。
3. 不. 该 P-值告诉你增量不是分派动物到处理组或对照组这一过程中的侥幸。该 P-值并不帮助你从给高剂量于老鼠到低剂量于人类的外推。
4. 模型在那里？为什么较低的工资是歧视的证据？(你可能不得不查一下经验、教育、生产率等等。)如果该专家坚持做检验，这些配对是非常相依的——比如，很可能有一名高报酬的男工程师在 16 个配对中出现。这种情况不同于抛硬币。

E 组. 第 609 页

1. 两样本 6——检验不适宜，因为这里不存在概率样本。不过与助教们相比学生们做得相当好。
2. 这里统计显著没多大意义，两颗内行星不构成取自内行星总体的容量为 2 的随机样本。它们是内行星。对外行星分析也类同。
3. 在这里显著性检验不适宜。这不是一个概率样本。
4. 问题有意义，因为我们在处理概率样本。但是，它不能在给定的信息基础上进行解答。这是一个群组样本，故简单随机样本的公式不适用：第 21 章第 4 节与第 22 章第 5 节。
5. 这样本是如此之大，因此任何差可能都是高度显著的
6. 显著性检验正在依据数据针对整个总体(“名流”们)进行。在这里盒子模型没多大意义。

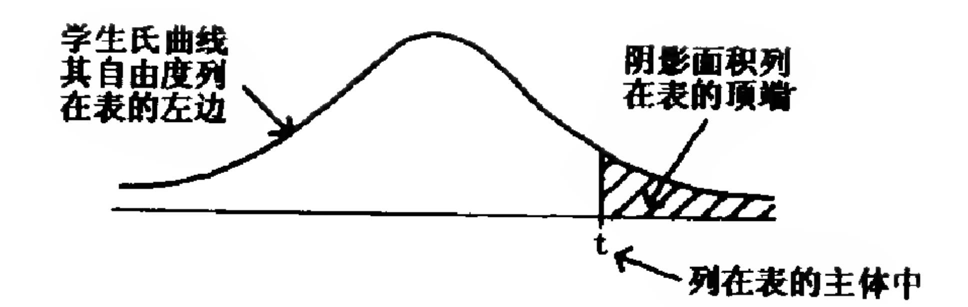
附表



正态表

Z	高	面积	Z	高	面积	Z	高	面积
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	1.98	98.57	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973
1.25	18.26	78.87	2.75	0.91	99.40	4.25	0.005	99.9979
1.30	17.14	80.64	2.80	0.79	99.49	4.30	0.004	99.9983
1.35	16.04	82.30	2.85	0.69	99.56	4.35	0.003	99.9986
1.40	14.97	83.85	2.90	0.60	99.63	4.40	0.002	99.9989
1.45	13.94	85.29	2.95	0.51	99.68	4.45	0.002	99.9991

t—表



自由度	25%	10%	5%	2.5%	1%	0.5%
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.69	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.60
5	0.73	1.48	2.02	2.57	3.36	4.03
6	0.72	1.44	1.94	2.45	3.14	3.71
7	0.71	1.41	1.89	2.36	3.00	3.50
8	0.71	1.40	1.86	2.31	2.90	3.36
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11
12	0.70	1.36	1.78	2.18	2.68	3.05
13	0.69	1.35	1.77	2.16	2.65	3.01
14	0.69	1.35	1.76	2.14	2.62	2.98
15	0.69	1.34	1.75	2.13	2.60	2.95
16	0.69	1.34	1.75	2.12	2.58	2.92
17	0.69	1.33	1.74	2.11	2.57	2.90
18	0.69	1.33	1.73	2.10	2.55	2.88
19	0.69	1.33	1.73	2.09	2.54	2.86
20	0.69	1.33	1.72	2.09	2.53	2.85
21	0.69	1.32	1.72	2.08	2.52	2.83
22	0.69	1.32	1.72	2.07	2.51	2.82
23	0.69	1.32	1.71	2.07	2.50	2.81
24	0.68	1.32	1.71	2.06	2.49	2.80
25	0.68	1.32	1.71	2.06	2.49	2.79

卡方表

自由度	99%	95%	90%	70%	50%	30%	10%	5%	1%
1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	4.66	6.57	7.79	10.82	13.34	16.22	21.06	23.69	29.14
15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	5.81.	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57

改编自 R. A. Fisher 爵士，Statistical Methods for Rcsearch Workers，(Edinburgh：Oliver & Boyd，1958)的第 134 页。

资料来源：摘自 R・A・Fishe 先生的《研究人员统计方法》。